Semi-supervised Infrared Small Target Detection with Thermodynamic-Inspired Uneven Perturbation and Confidence Adaptation

Mingjin Zhang[1], Wenteng Shang[1], Fei Gao[1*], Qiming Zhang[2], FengQin Lu[1], Jing Zhang[3]

[1]Xidian University, Xi'an 710126, China
[2]School of Computer Science, The University of Sydney, NSW 2006, Australia
[3]School of Computer Science, Wuhan University, Wuhan 430072, China
{mjinzhang, 21011210140, fgao}@xidian.edu.cn, qzha2506@uni.sydney.edu.au, jingzhang.cv@gmail.com

Abstract

Single-frame Infrared Small Target (SIRST) detection has made significant advancements, but it still faces challenges due to limited labeled data and the foreground-background class imbalance. To address these issues, we introduce a novel *Semi-Supervised SIRST Detection* (S^3D) pipeline in this paper. First, drawing inspiration from thermodynamics, we propose augmenting infrared images using both *chromatically* and *spatially* uneven perturbations. This dual-stream perturbation enhances the diversity and balance of infrared samples, contributing to the robustness of detection models. Additionally, we develop a confidence-adaptive matching method to maintain weighted consistency among *perturbed* unlabeled samples. Second, to tackle class imbalance in labeled data, we compel the model to generate discriminative predictions for challenging, misclassified examples while down-weighting well-classified examples. We achieve this by modifying the standard cross-entropy loss to *squeeze* the detector and truncating the loss on well-classified examples. Our innovative *Truncated Squeeze* loss focuses on learning discriminative representations for difficult cases and prevents over-optimization for simpler ones. To assess the effectiveness of the perturbation techniques and loss functions, we apply them to various SIRST detectors and conduct extensive experiments on two benchmarks. Notably, our methods consistently and significantly improve accuracy. Remarkably, it achieves over 98% performance of the state-of-the-art fully-supervised method using only 1/8 of the labeled samples.

Introduction

Single-frame infrared small target (SIRST) detection is a critical task in infrared (IR) search and tracking. It has a wide range of diverse applications, such as maritime search and rescue, and agricultural yield prediction (Zhang et al. 2024a; Chen et al. 2024; Chen, Wu, and Wang 2019). SIRST detection relies solely on spatial information from a single image, which offers advantages like easy deployment and real-time inference. These benefits make SIRST detection highly appealing for identifying fast-moving, isolated IR small targets. Traditional SIRST methods (Bai and Zhou 2010; Zhang and Peng 2019), including filtering-based, local contrast-based, and low-rank-based approaches, often require complex feature design and parameter tuning, resulting in limited accuracy. In recent years, due to the significant success of deep learning techniques (Ren et al. 2015; Redmon and Farhadi 2018), SIRST detection has made considerable progress (Zhang et al. 2024b, 2022c; Li et al. 2023). However, existing methods primarily rely on fully-supervised learning. Their performance is heavily dependent on the quality and quantity of labeled data. Unfortunately, due to the expensive data collection of the SIRST task, its development is facing great challenges such as scarce labeled data, foreground-background class imbalance, and label noise in manual annotations. To address these issues, we propose to resort to the semi-supervised learning technique to enhance SIRST detection performance.

Recently, semi-supervised learning methods (Berthelot et al. 2020; Li, Socher, and Hoi 2019; Liu et al. 2020), particularly those employing consistency regularization, have demonstrated considerable potential in the field of computer vision. A key challenge in these methods is designing effective perturbation techniques to enhance generalization capabilities. However, existing data perturbation techniques (Wang et al. 2023; Yang et al. 2023; Ghosh and Thiery 2020) primarily cater to normal images and are not specifically tailored for IR data. Given the differences in imaging principles between IR images and photographs, it is essential to explore specialized data perturbation methods for IR data.

From a thermodynamic perspective, an imaging area can be considered a local thermodynamic system in thermal equilibrium, exhibiting stable internal energy changes. In practice, various factors, such as uneven radiation, motion blur, and shooting angle, may cause spatially and chromatically uneven distortion in an acquired IR image. We leverage the thermal distribution contrast between turbulent and equilibrium states, denoted as ΔQ (Fig. 1). As the thermodynamic system approaches equilibrium, heat exchange becomes directionless, and ΔQ diminishes. This enables us to create uneven chromaticity augmentation by aligning the temperature field function with the energy distribution changes in the IR image's microelements. To account for factors affecting IR target positions, such as shooting angle and motion blur, we develop a spatially uneven perturbation. This dual-stream uneven perturbation enhances both the diversity and balance of infrared samples, ultimately benefiting the robustness of detection models.

*Corresponding author.

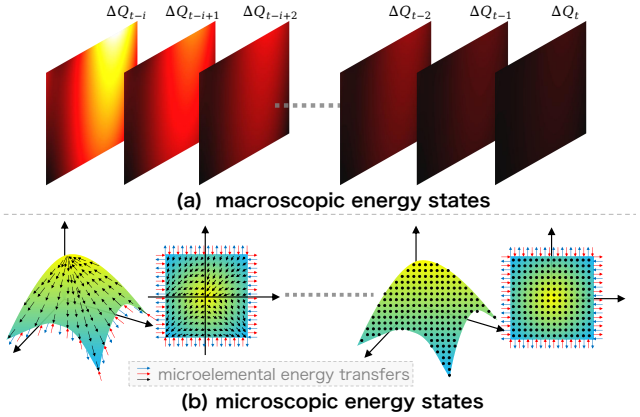

Figure 1: Illustration of the energy distribution changes (ΔQ), in a thermodynamic system, approaching a steady state. It involves changes in both macroscopic (a) and microscopic (b) energy states. The arrows in (b) indicate the direction of microelemental energy transfers.

In the supervised learning context, addressing the foreground-background class imbalance in labeled data is critical. Previous imbalanced learning methods, such as Focal Loss (Lin et al. 2017), primarily focus on down-weighting well-classified examples to boost the relative importance of misclassified examples. However, these methods actually reduce the absolute loss on misclassified examples. In this paper, we propose a dual approach: compelling the model to generate discriminative predictions on challenging, misclassified examples, and down-weighting well-classified examples. To achieve this, we modify the standard cross-entropy loss to "squeeze" the detector and truncate the loss on well-classified examples. Our novel Truncated Squeeze (TS) loss emphasizes learning discriminative representations for difficult cases and prevents over-optimization for simpler ones. In the unsupervised learning context, previous methods (Yang et al. 2023) typically force the strongly perturbed predictions to align with the weakly perturbed ones. However, in our scenario, the chromatic perturbation (strong) often causes significant deviation in an infrared image. In other words, the strongly perturbed views might vary significantly from each other and from the weakly perturbed view. Therefore, we design a weighted strong-to-strong consistency loss to adaptively modulate the contribution of unlabeled data to the optimization process.

Building on prior discussions, we propose a novel Semi-Supervised SIRST Detection (S3D) pipeline. To verify the effectiveness of our perturbation techniques and loss functions, we apply them to various SIRST detectors and conduct extensive experiments on two benchmark datasets. Notably, our approach consistently enhances accuracy and achieves over 98% of the performance of state-of-the-art fully-supervised ones with only 1/8 of the labeled samples.

In summary, our contributions are mainly four-fold:

- We propose an infrared imaging-inspired uneven perturbation, which significantly enhances accuracy and effectively complements commonly used uniform data augmentation techniques.
- We introduce a novel *Truncated* Squeeze (TS) loss to address the class imbalance issue. The TS loss has been proven to consistently enhance performance across various network architectures.
- We devise a weighted strong-to-strong consistency loss, which adaptively modulates the contribution of unlabeled data based on the confidence of predictions.
- Our final *Semi-Supervised SIRST Detection* (S^3D) method achieves remarkable performance on challenging benchmarks and approaches the fully-supervised performance while using only 1/8 of the labeled samples.

Related Works

SIRST Detection

Researchers have proposed various methods over time, including traditional methods (filtering-based methods (Bai and Zhou 2010; Deshpande et al. 1999), local comparison-based methods (Chen et al. 2013; Han et al. 2020; Hou and Zhang 2007; Han et al. 2014), low-rank-based methods (Dai and Wu 2017; Gao et al. 2013; Zhang and Peng 2019; Zhang et al. 2018)), which is similar to the idea of low-rank representations (Zhang et al. 2019a) and probabilistic graph modelling (Zhang et al. 2019b) in face recognition, and deep learning-based methods (Zhang et al. 2022b, 2023, 2024b).

Traditional methods require elaborate mathematical modeling and hyperparameter tuning, while deep learning-based methods learn complex nonlinear mappings in a data-driven manner, generalizing well to IR datasets with diverse characteristics and showing promising performance. Dai *et al.* design an ACMNet (Dai et al. 2021a) to facilitate the interaction between high-level and low-level information. Zhang *et al.* propose an ISNet (Zhang et al. 2022c) to leverage image edge information with Taylor finite-difference equations. Additionally, they innovate with an RKFormer (Zhang et al. 2022a), incorporating a transformer-based approach to optimize the self-attention module for full-feature detection. Li *et al.* present a DNANet (Li et al. 2023) to facilitate the multi-scale fusion of IR small target features. For exploring lightweight network architectures, Zhang *et al.* develop an IRPruneDet (Zhang et al. 2024a) through the wavelet structure-regularized soft channel pruning; Li *et al.* design an SPMix-Q (Li et al. 2024) with the low-bit quantization.

Current methods improve IR target detection through network design but rely on full supervision, which depends on high-quality and abundant datasets. However, limited single-frame IR data and inconsistent labeling hinder SIRST detection research. To overcome this, we pioneer the application of semi-supervised learning in SIRST detection, striving to lessen dependence on labeled data and augment the IR dataset effectively.

Semi-Supervised Semantic Segmentation

The core solution to the semi-supervised problem lies in how to design a training strategy to utilize a large number of unlabeled samples to align labeled and unlabeled features to

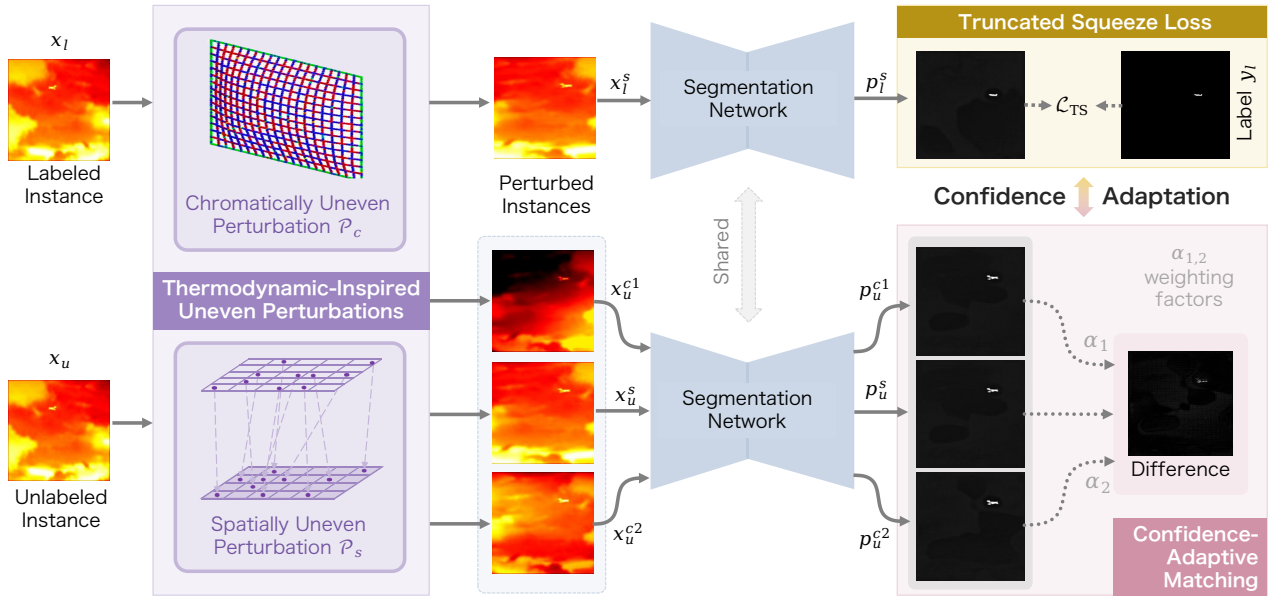

Figure 2: Overview of the proposed *semi-supervised SIRST detection* (S^{3}D) pipeline. DNANet (Li et al. 2023) is adopted as the default segmentation network. The input labeled/unlabeled data are spatially or chromatically perturbed. The *Truncated Squeeze* (TS) loss on labeled data, coupled with the weighted consistency on unlabeled data, are used to optimize the network.

enhance the generalization (Na et al. 2023; Xie, Sun, and Li 2023; Chen et al. 2020). Recent methods typically fall into two categories (Hu et al. 2021): entropy minimization (Jin, Wang, and Lin 2022) and consistency regularization (Xu et al. 2022). Entropy minimization aims to increase prediction confidence, with self-training as the primary method, where the network uses its own predictions as pseudo-labels for supervision. Despite filtering for accuracy, pseudo-labels can still suffer from confirmation bias. This is repeated until the network outputs low entropy predictions for unlabeled samples (Sun et al. 2023). Yang *et al.* develop ST++ (Yang et al. 2022), showcasing that a straightforward iterative self-training approach yields good performance.

Another class of methods relies on the smoothing and clustering assumptions, positing that the network generates similar predictions for identical samples with varying perturbations. Chen *et al.* propose a cross pseudo supervision strategy (Chen et al. 2021) to employ two independent networks to supervise each other. FixMatch (Sohn et al. 2020) injects both strongly and weakly enhanced images and uses the weakly enhanced predictions to supervise the strongly enhanced ones. FlexMatch (Zhang et al. 2021) and FreeMatch (Wang et al. 2023) introduce curriculum learning methods without adding extra parameters, adapting to different training states and class difficulties. Recently, Unimatch (Yang et al. 2023) highlights the importance of increasing the image-level perturbation space and designing effective strong augmentations. This insight forms the theoretical basis for applying image-level data augmentation to address spatial distortion challenges in IR imaging for our SIRST detection task.

Methodology

Let $\mathcal{D}_l = \{(x_l^i, y_l^i)\}_{i=1}^{N_l}$ and $\mathcal{D}_u = \{x_u^j\}_{j=1}^{N_u}$ denote the set of labeled and unlabeled data, respectively. The entire dataset is $\mathcal{D} = \mathcal{D}_l \cup \mathcal{D}_u$. Our goal is to learn a detector $\phi(x)$ to predict the segmentation mask of targets, y, in an infrared image x. Fig. 2 illustrates the proposed *semi-supervised SIRST detection* (S^{3}D) pipeline. We use DNANet (Li et al. 2023) as our default segmentation network. Note that, as will be demonstrated in the experimental section, we can use arbitrary segmentation networks alternatively. To tackle the challenge of scarce labeled data and class imbalance, we propose adding uneven perturbations to the data, and using a supervised balance loss and an unsupervised consistency loss, to optimize the network. Details will be introduced below.

Thermodynamic-Inspired Uneven Perturbations

Chromatically Uneven Perturbation (*Strong*). IR cameras typically capture targets with spatial distortions, due to the uneven radiation, target motion blur, shooting angle interplay, or background thermal noise. To simulate such chromatic distortions, we propose a chromatically uneven perturbation method, based on the intuitive mapping between IR images and thermal radiation systems. In a thermal radiation system, the energy difference ΔQ between disordered and equilibrium states is continuous (Fig. 1). Through internal thermal interactions and boundary energy exchanges, the system moves towards equilibrium, reducing ΔQ to zero. By utilizing the principles of energy continuity and unevenity, we simulate a random energy distribution in thermal radiation systems, by using a smooth two-dimensional uneven stochastic chromatic map.

Specifically, to simplify generating a smooth energy dis-

tribution, we decompose this two-dimensional problem into two one-dimensional problems and generate five random points $\{A, B, C, D, E\}$.

In the vertical dimension, we assume the vertical thermal distribution is highly correlated, leading to smooth variation. Inspired by the concept of temperature field in thermodynamics, we design a temperature field function $T(y)$ to describe the temperature of each point (or pixel) in an IR image, where y signifies the vertical position. According to the Kirchhoff's law (Kirchhoff 1860), smooth variations in the vertical temperature field value result in smooth changes in the energy state. Such smoothness implies continuous changes in neighboring pixels' grayscale values, in an IR image. To achieve this, the function $T(y)$ must have both bounded values and derivatives (*i.e.*, the rate of temperature change), across the entire image.

In the horizontal dimension, according to the Stefan-Boltzmann law (Paul et al. 2015), the total energy of blackbody radiation is proportional to the fourth power of its absolute temperature, *i.e.* $Q = \int_0^\infty Q_\lambda(T)d\lambda = \sigma T^4$. Considering this, we convert the previously generated five random points, in the horizontal direction, to a new set of five random points in the vertical direction, through a linear function. This allows these random endpoints to smoothly transform to the next position. We assume $f(x)$ is the indefinite integral of a quadratic function:

$$f(x) = \varepsilon \int ax^3 + bx^2 + cx + d, \tag{1}$$

where $f(x)$ represents the energy state of the pixel at position x. In our setup, we employ a uniform step size for simplicity. When the initial random five points move to their endpoints, a random map featuring a smooth energy distribution is generated. These maps exhibit both horizontal and vertical smoothness and randomness. Such chromatic uneven perturbation, randomly enhances or diminishes the foreground-background contrast, simulating heat variations. The concise pseudo-code of our chromatic perturbation is shown in Algorithm 1.

Spatially Uneven Perturbation (*Weak*). IR images are affected by external factors such as imaging angle, distance, and atmospheric radiation disturbances, leading to inherent randomness and distortion in the position of targets and backgrounds. Besides, internal factors like the spatially uneven response of individual detectors in the infrared sensor can cause fixed pattern noise (Guan et al. 2020), resulting in blurring and uneven degradation. In practice, we remap each pixel's position (x, y). Let $g(x, y)$ be the target image, $f(x, y)$ the source image, and $h(x, y)$ the mapping function:

$$g(x, y) = f(h(x, y)). \tag{2}$$

The effect of the deformation enhancement depends entirely on the design of $h(x, y)$. To achieve multiple stochastic effects of smooth stretching and shrinking simultaneously in the same dimension, we design a sine function. The sine function's amplitude is based on 60 and randomly varies by ±15 pixels. For mapping in the x-direction, the formula for $h(x, y)$ is:

$$h(x, y) = (x + \delta * \sin(2 * \pi * x/k), y), \tag{3}$$

Algorithm 1: Chromatically uneven perturbation $\mathcal{P}_c$

Input: Five randomly sampled pixels $A_1(x_a, y_a, z_{a1})$, $B_1(x_b, y_b, z_{b1})$, $C_1(x_c, y_c, z_{c1})$, $D_1(x_d, y_d, z_{d1})$, $E_1(x_e, y_e, z_{e1})$, and their corresponding augmented pixels $A_2(x_a, y_a, z_{a2})$, $B_2(x_b, y_b, z_{b2})$, $C_2(x_c, y_c, z_{c2})$, $D_2(x_d, y_d, z_{d2})$, $E_2(x_e, y_e, z_{e2})$.
Output: Chromaticity offset image M.

for $i \leftarrow 0$ to $N-1$ **do**
  $M_i \leftarrow$: define M_i as the set of sets of image positions as well as luminance sets
  for $y \leftarrow 0$ to $h-1$ **do**
    for $P \in \{A, B, C, D, E\}, j \in \{a, b, c, d, e\}$ **do**
      $P \leftarrow (x_j, z_j)$;
      $z_j \leftarrow z_{j1} + (z_{j2} - z_{j1}) \times \frac{y}{h-1}$;
    end for
    Fit the quadratic function (1) through the set P;
    for $x \leftarrow 0$ to $w-1$ **do**
      $z \leftarrow f(x)$;
      Add the augemented pixels to the set M_i: $M_i \leftarrow \{(x, y), z\}$;
    end for
  end for
  Generate chromaticity offset image: $M \leftarrow M \cup M_i$;
end for

where k is randomly generated within a set interval. Since coordinate points are discrete, this function generates a discrete position map. The mapping in the y-direction follows a similar approach.

Confidence-Adaptive Matching Consistency

Inspired by consistency-regularized semi-supervised methods like FixMatch (Sohn et al. 2020) and UniMatch (Yang et al. 2023), we use the confidence-adaptive matching consistency on unlabeled data, to optimize the detector besides the supervised branch. We randomly perform chromatic ($\mathcal{P}_c$) and spatial ($\mathcal{P}_s$) perturbation on an input $x_u^j \in \mathcal{D}_u$; and then estimate the corresponding semantic labels by:

$$p_u^s = \phi(\mathcal{P}_s(x_u^j)), \quad p_u^c = \phi(\mathcal{P}_c(\mathcal{P}_s(x_u^j))), \tag{4}$$

where $\phi(\cdot)$ denotes the segmentation network. In practice, we have one spatially perturbed estimate p_u^s, and two chromatically perturbed estimates p_u^{c1} and p_u^{c2}.

Recall that the chromatic deviation produces strong perturbations on the original targets; and the spatial deviation only produces weak perturbations in shape. Therefore, we can assume the weakly perturbed estimation p_u^s as confident, and determine optimization weights, by comparing the consistency between predictions $\{p_u^s, p_u^{c1}, p_u^{c2}\}$. The confidence-adaptive consistency loss is formulated as:

$$\ell_{con}(x_u^j) = \frac{1}{2}\alpha_1 * D(\mathrm{sg}[p_u^{c1}], p_u^{c2}) + \frac{1}{2}\alpha_2 * D(p_u^{c1}, \mathrm{sg}[p_u^{c2}]), \tag{5}$$

$$\alpha_k = \frac{D(p_u^s, p_u^{ck})}{D(p_u^s, p_u^{c1}) + D(p_u^s, p_u^{c2})}, k = 1, 2. \tag{6}$$

where $\mathrm{sg}[\cdot]$ denotes the stop-gradient operation; $D(\cdot)$ is the L1 distance between two predictions; $\alpha_k (k = 1, 2)$ is the

consistency-based weighting factor. Initially, the chromatically perturbed images should have the same semantic mask, and $D(p_u^{c1}, p_u^{c2})$ should be minimized. In UniMatch, the strongly perturbed predictions are forced to approach the weakly perturbed ones, with a confidence threshold. However, in our scenario, the chromatic perturbation typically causes great deviation in an infrared image; p_u^{c1} and p_u^{c2} might diverse greatly from each other, and from p_u^s. Thus, we use p_u^s as an anchor, to evaluate the confidences of p_u^{c1} and p_u^{c2}. The contribution of unlabeled data to the optimization process, is adaptively relaxed. The total consistency loss across all the unlabeled data is denoted by:

$$\mathcal{L}_{con} = \frac{1}{N_u} \sum_{j=1}^{N_u} \ell_{con}(x_u^j). \tag{7}$$

Imbalanced Learning via Truncated Squeeze Loss

In the SIRST detection task, the target typically occupies a small portion of the image, causing a notable imbalance between positive and negative samples. Such imbalance, coupled with limited labeled data and correlated noise, poses a significant challenge in learning discriminative representations. To tackle this challenge, we propose a *Truncated Squeeze Loss* (TSL) function that focuses on optimizing difficult samples. For a pixel in arbitrary position, the TS loss is calculated by:

$$\ell_{\mathrm{TS}}(p_t) = w(p_t) \cdot \log p_t \tag{8}$$

with

$$\begin{aligned} w(p_t) &= \mathbb{1}(p_t < \eta_{C^*}) \cdot e^{1-p_t} \\ &= \begin{cases} e^{1-p_t}, & \text{if } p_t < \eta_{c^*} \\ 0, & \text{otherwise} \end{cases} \end{aligned} \tag{9}$$

where p_t is the predicted probability belonging to the real category c^* (Lin et al. 2017). η_{c^*} is the corresponding tunable threshold. The normalized total loss, across all the labeled examples, is computed by:

$$\mathcal{L}_{\mathrm{TS}} = \sum_{i=1}^{N_l} \ell(p_t^i) / \sum_{i=1}^{N_l} w(p_t^i), \tag{10}$$

where i is the instance index. In our experiments, we set different thresholds for positive and negative examples, *i.e.* $\eta_{pos} = 0.99$ and $\eta_{neg} = 0.9$ sequentially. By using such asymmetrical thresholds, the learning process emphasizes more on well classifying positive examples (targets/foreground), than negative ones (background). In this way, $\mathcal{L}_{\mathrm{TS}}$ counteracts the class imbalance to some extent.

Fig. 3 shows the curves of the BCE loss, Focal loss (FL) (Lin et al. 2017), and TS loss. The BCE loss is formulated as $\ell_{\mathrm{CE}}(p_t) = -\log(p_t)$; the FL loss is $\ell_{\mathrm{FL}}(p_t) = -(1-p_t)^{\gamma}\log(p_t)$. Compared to BCE and FL, our TS loss significantly increases the loss on misclassified examples. Besides, our TS loss zero sets the loss on well-classified examples. Thus, the TS loss forces the model to produce discriminative predictions, and forbids over-optimization to easy examples. In addition, there might be noise in manually labeled annotations, especially around the boundary regions. Overfitting to such label noise, generally decrease the generalization capacity of the learned detectors (Li et al. 2022).

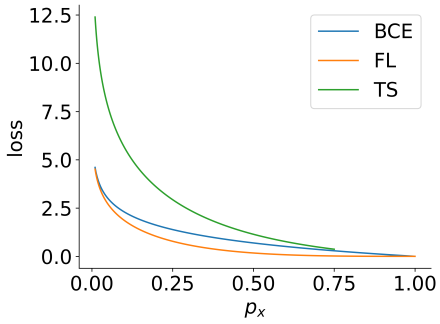

Figure 3: Curves of the BCE, FL ($\gamma = 2$), and TS losses.

In our TS loss, the truncation of loss, with a relatively low threshold, mitigates the influence of such label noise.

Total Loss. The overall objective is a combination of supervised loss and unsupervised loss, *i.e.*

$$\mathcal{L}_{all} = \mathcal{L}_{\mathrm{TS}} + \lambda \mathcal{L}_{con}, \tag{11}$$

where $\lambda = 0.5$ is the weighting parameter.

Experiments

Experimental Settings

Datasets. We conduct extensive experiments on our S^{3}D on the publicly available NUDT-SIRST (Li et al. 2023) and NUAA-SIRST (Dai et al. 2021b). NUAA-SIRST has 427 IR images with backgrounds such as clouds, cities, and oceans. NUDT-SIRST contains the most diverse background categories, including clouds, cities, oceans, and bright lights. For all datasets, we use 20% for testing and 80% for training.

Evaluation Metrics. We utilize two popular pixel-level evaluation metrics, including the *Intersection over Union* (IoU) and the *False alarm rate* (F_a), to measure the performance of IR detection. Both IoU and F_a are sensitive to the prediction of small targets.

Implementation Details. To facilitate training, all images are resized to 256×256. Each batch contains 4 labeled and 2 unlabeled images. The initial learning rate is set to 0.0001, employing an AdamW optimizer. Parameters remain consistent for different datasets. Original images are horizontally or vertically flapped with a 0.2 probability, while the uneven spatial perturbation is performed with a 0.6 probability.

Quantitative Comparisons with SOTA Methods

Comparison with Semi-supervised Methods. We apply the proposed techniques to ACMNet (Dai et al. 2021a), DNANet (Li et al. 2023), ResUNet (Ronneberger, Fischer, and Brox 2015)), and train them in the semi-supervised learning manner. The corresponding model variants are denoted by †. We use ResUNet as our default segmentation network, thus ResUNet† is our official model, *i.e.* S^{3}D. As shown in Table. 1, our method S^{3}D consistently shows significant superiority, over current semi-supervised frameworks, i.e. ST++ and CPS. Specifically, S^{3}D achieves 10-20% absolute superiority in IoU, and 40-100 ($\times 10^{-6}$) absolute decrement in F_a, on NUDT-SIRST. Moreover, the proposed semi-supervised techniques, consistently boosts the

	NUDT-SIRST						NUAA-SIRST					
	1/32		1/16		1/8		1/32		1/16		1/8	
	IoU↑	F_a↓	IoU↑	F_a↓	IoU↑	F_a↓	IoU↑	F_a↓	IoU↑	F_a↓	IoU↑	F_a↓
ST++ (Yang et al. 2022)	49.88	110.28	64.91	93.29	80.72	17.19	52.19	33.12	58.62	21.12	62.14	17.19
CPS (Chen et al. 2021)	50.18	126.28	61.93	94.12	76.53	57.12	52.49	29.17	57.38	24.76	61.94	18.92
ACMNet (Dai et al. 2021a)	58.42	48.83	64.94	30.40	68.80	32.24	44.19	90.12	50.28	62.18	57.79	47.16
ACMNet†	61.04	63.57	69.19	23.67	73.21	23.38	48.19	76.12	55.28	60.29	60.29	27.19
relative (%)	4.48↑	30.19↑	6.54↑	22.14↓	6.41↑	27.48↓	9.05↑	15.53↓	9.94↑	3.04↓	4.33↑	42.35↓
ResUNet (Ronneberger, Fischer, and Brox 2015)	46.35	160.25	57.04	79.24	76.15	47.12	48.45	60.27	53.58	30.14	58.94	23.19
ResUNet†	**77.61**	**21.04**	82.14	**9.29**	85.07	11.60	**61.58**	**21.94**	62.48	**13.01**	66.40	**7.24**
relative (%)	67.44↑	86.87↓	44.00↑	88.28↓	11.71↑	75.38↓	27.10↑	63.60↓	16.61↑	56.83↓	12.66↑	68.78↓
DNANet (Li et al. 2023)	69.39	65.27	76.29	22.10	84.30	17.89	52.14	75.28	67.07	29.18	70.06	18.21
DNANet† (S^3D)	76.58	53.15	**84.59**	20.90	**90.12**	**6.85**	56.79	70.12	**69.89**	20.12	**73.12**	10.28
relative (%)	10.36↑	18.57↓	10.88↑	5.43↓	6.90↑	61.71↓	8.92↑	6.85↓	4.20↑	31.05↓	4.37↑	43.55↓

Table 1: Comparison with SOTA methods in IoU(%), P_d(%), $F_a(10^{-6})$, on NUDT-SIRST and NUAA-SIRST. ST++ and CPS are semi-supervised methods; ACMNet, ResUNet, and DNANet are supervised methods, and are trained with our $\mathcal{L}_{\mathrm{TS}}$ on the subset of labeled data. † indicates the corresponding *semi-supervised* model variant, with our proposed techniques.

	fully-supervised & trained with the boosted TS loss														*semi-sup.* (1/8)	
	MDvsFA		ALCNet		ISNet		GCI-Net		ACMNet		ResUNet		DNANet		S^3D (Ours)	
	IoU↑	F_a↓	IoU↑	F_a↓	IoU↑	F_a↓	IoU↑	F_a↓	IoU↑	F_a↓	IoU↑	F_a↓	IoU↑	F_a↓	IoU↑	F_a↓
NUDT-SIRST	45.38	200.71	61.78	36.36	71.27	96.84	72.33	79.95	75.19	18.18	90.28	5.89	**91.89**	**1.79**	90.12	6.85
NUAA-SIRST	61.77	64.90	67.91	37.04	72.04	42.46	74.31	20.21	64.92	12.76	77.18	**9.23**	**79.86**	15.50	73.12	10.28

Table 2: Comparison with *fully-supervised* SIRST methods, which are trained with $\mathcal{L}_{\mathrm{TS}}$ and all labeled data.

performance on all networks (e.g. ACMNet† vs. ACMNet), across almost all the settings. Remarkably, on ResUNet, the proposed techniques boost IoU by 44-67% and decrease F_a by over 80% relatively, while using 1/32 and 1/16 labeled data on NUDA-SIRST. Note that ACMNet, DNANet, and ResUNet here are trained with our $\mathcal{L}_{\mathrm{TS}}$, and are better than their initial versions (Table 4). Finally, the proposed techniques lead to larger performance improvement, as less labeled data are available. Such phenomenon implies great potentials, of our method, in few-shot learning tasks.

Comparison with Fully-supervised Methods. We compare with SOTA semi-supervised methods (CPS (Chen et al. 2021) and ST++ (Yang et al. 2022)), and several supervised methods (MDvsFA (Wang, Zhou, and Wang 2019), ALCNet (Dai et al. 2021b), ISNet (Zhang et al. 2022c), GCI-Net (Zhang et al. 2024b), ACMNet, DNANet, and ResUNet. We additionally illustrate the performance of fully-supervised SIRST detection methods in Table 2. We train these methods with our $\mathcal{L}_{\mathrm{TS}}$ and the standard training data, to explore the upper bound of performance. By using 1/8 labeled data, our S^3D surpasses most fully-supervised methods, and achieves competitive results with the best performance, on both datasets. Remarkably, S^3D (1/8) surpasses all these methods in terms of F_a on NUAA-SIRST. Such low F_a implies that, our methods prevent the background from dominating the optimization process. Besides, the IoU of S^3D (1/8) approaches 98% of fully-supervised DNANet on NUDT-SIRST, and 92% on NUAA-SIRST.

Ablation Studies

Impact of Uneven Perturbation. We first analyze the impact of our dual-stream uneven perturbation on the performance. As shown in Table 3, either the chromatic perturbation $\mathcal{P}_c$ or the spatial perturbation $\mathcal{P}_s$, boosts the performance, in both IoU and F_a, under all settings. Besides, using both $\mathcal{P}_c$ and $\mathcal{P}_s$ consistently leads to the best performance. Such observations demonstrate the contribution of the proposed uneven perturbation to SIRST detection.

Complement to Uniform Augmentation. To verify the superiority of the proposed two-stream perturbations, we compare with common *uniform* augmentation techniques, including (1) geometric transformation $\mathcal{A}_g$, e.g. random horizontal flipping, vertical flipping, and random scaling; (2) luminance transformation $\mathcal{A}_l$, e.g. Gaussian blur, random contrast, etc.; (3) additive Gaussian noise $\mathcal{N}$. As shown in Table 3, both $\mathcal{A}_g$ and $\mathcal{A}_l$ significantly boosts the IoU. While Gaussian noise $\mathcal{N}$ slight decreases the performance. In contrast, $\mathcal{P}_c$ further boosts the performance. Such comparison demonstrates the superiority and of our chromatic perturbation method. Besides, using both $\mathcal{P}_c$ and $\mathcal{P}_s$ leads to the best performance consistently. Such observations demonstrate that our uneven perturbations well complement to uniform augmentation methods.

Impact of Loss Functions. Finally, we verify the superiority of our TS loss $\mathcal{L}_{\mathrm{TS}}$ and consistency loss $\mathcal{L}_{con}$. We train a detector with the BCE loss, IoU loss, or Focal loss,

Perturb.		1/32		1/16		1/8		1	
$\mathcal{P}_c$	$\mathcal{P}_s$	IoU↑	F_a↓	IoU↑	F_a↓	IoU↑	F_a↓	IoU↑	F_a↓
-	-	72.28	45.03	76.68	35.30	79.61	30.29	87.51	10.93
✓	-	74.37	32.10	79.79	20.67	83.25	19.87	89.78	6.69
-	✓	73.83	30.52	78.60	18.02	83.00	20.27	89.37	6.35
✓	✓	**77.61**	**21.94**	**82.14**	**9.27**	**85.07**	**11.60**	**90.28**	**5.89**

(IoU↑)	1/32	1/16	1/8
N/A	58.67	61.25	64.81
$\mathcal{A}_g$+$\mathcal{A}_l$	72.28	76.68	79.61
$\mathcal{A}_g$+$\mathcal{A}_l$+$\mathcal{N}$	71.35	76.28	78.38
$\mathcal{A}_g$+$\mathcal{A}_l$+$\mathcal{P}_c$	74.37	79.79	83.25
$\mathcal{A}_g$+$\mathcal{A}_l$+$\mathcal{P}_c$+$\mathcal{P}_s$	**77.61**	**82.14**	**85.07**

Table 3: Ablation study of data perturbation, with semi-supervised ACMNet on NUDT-SIRST.

(*Semi.*)	1/32		1/16		1/8	
ResUNet	IoU↑	F_a↓	IoU↑	F_a↓	IoU↑	F_a↓
$\mathcal{L}_{\text{BCE}}$	36.01	126.39	46.85	68.92	55.81	62.82
$\mathcal{L}_{\text{IoU}}$	71.37	62.47	74.99	43.18	78.21	16.58
$\mathcal{L}_{\text{FL}}$ (Lin et al. 2017)	72.76	**21.13**	75.31	37.95	81.21	11.80
$\mathcal{L}_{\text{TS}}$(η_{c^*}=1)	76.35	27.75	79.09	30.81	82.84	20.33
$\mathcal{L}_{\text{TS}}$	**77.61**	21.94	**82.14**	**9.27**	**85.07**	**11.60**
$\mathcal{L}_{\text{UniMatch}}$ (Yang et al. 2023)	77.13	33.28	81.12	11.74	83.97	13.18

(*Sup.*)	1/32		1/16		1/8	
	IoU↑	F_a↓	IoU↑	F_a↓	IoU↑	F_a↓
ACMNet						
$\mathcal{L}_{\text{IoU}}$	54.29	46.29	63.12	44.57	65.24	38.21
$\mathcal{L}_{\text{TS}}$	**58.42**	**48.83**	**64.94**	**30.40**	**68.80**	**32.24**
DNANet						
$\mathcal{L}_{\text{IoU}}$	63.57	88.23	73.19	23.89	76.15	20.56
$\mathcal{L}_{\text{TS}}$	**69.39**	**65.27**	**76.29**	**22.10**	**84.30**	**17.89**

Table 4: Ablation study of loss functions, with diverse networks, on NUDT-SIRST.

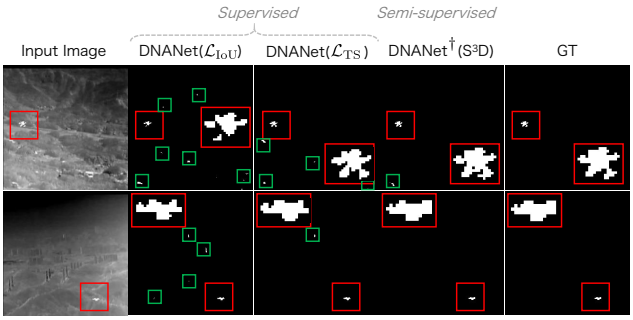

Figure 4: SIRST detection results: red boxes indicate correctly detected targets, and green boxes indicate false detections.

instead of $\mathcal{L}_{\text{TS}}$. As shown in Table 4, $\mathcal{L}_{\text{TS}}$ leads to the best performance on ResUNet in general. Besides, using $\mathcal{L}_{\text{TS}}$ significantly boost the performance of both ACMNet and DNANet, compared to $\mathcal{L}_{\text{IoU}}$. In addition, if we remove the truncation operation in $\mathcal{L}_{\text{TS}}$ (denoted by η_{c^*}=1), the performance gain (especially in F_a) reduces. Such diminution implies the significant role of truncation in balancing foregrounds and backgrounds. Finally, if we replace $\mathcal{L}_{con}$ with UniMatch (Yang et al. 2023), the performance consistently decreases. Such comparison meets our motivation of designing the weighted strong-to-strong consistency loss.

Qualitative Analysis. Fig. 4 illustrates the detection results by different model variants of DNANet, by using 1/32 labeled data on NUDT-SIRST. Obviously, the model learned with $\mathcal{L}_{\text{IoU}}$ produces many false positive predictions. In contrast, our $\mathcal{L}_{\text{TS}}$ significantly reduces the number of false positive examples. This observation is consistent with the quantitative metrics in Table 4. Finally, with all the proposed semi-supervised learning techniques, our full model, i.e. S^{3}D, produces the best predictions, with the fewest false positive examples. Besides, S^{3}D predicts the most accurate boundaries of targets. Such observations qualitatively demonstrate the superiority of the proposed techniques, in solving the scarcity of labeled data and the challenge of class imbalance.

Conclusion

We propose the first semi-supervised method for SIRST detection. Our S^{3}D framework includes three plug-and-play modules: uneven perturbation, truncated squeeze loss, and weighted consistency loss. Uneven perturbation, inspired by infrared imaging, expands IR datasets and improves network generalization. The loss functions tackle target-background imbalance and address labeling quality issues. Experiments confirm the effectiveness of these modules across various networks and datasets. Future work will explore applying these methods to other imbalanced detection tasks.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 92470108, Grant 62272363; in part by the Young Elite Scientists Sponsorship Program by China Association for Science and Technology (CAST) under Grant 2021QNRC001; in part by the Joint Laboratory for Innovation in Satellite-Borne Computers and Electronics Technology Open Fund 2023 under Grant 2024KFKT001-1; in part by the Proof of Concept Foundation of Xidian University Hangzhou Institute of Technology under Grant No. GNYZ2023YL0301; and in part by the Fundamental Research Funds for the Central Universities under Grant No. ZYTS24012.

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.

Berthelot, D.; Carlini, N.; Cubuk, E. D.; Kurakin, A.; Sohn, K.; Zhang, H.; and Raffel, C. 2020. ReMixMatch: Semi-Supervised Learning with Distribution Matching and Augmentation Anchoring. In *International Conference on Learning Representations*.

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.

Chen, Q.; Wu, C.; and Wang, Y. 2019. Robust principal component analysis-based infrared small target detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 9925–9926.

Chen, T.; Tan, Z.; Chu, Q.; Wu, Y.; Liu, B.; and Yu, N. 2024. TCI-Former: Thermal Conduction-Inspired Transformer for Infrared Small Target Detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 1201–1209.

Chen, X.; Yuan, Y.; Zeng, G.; and Wang, J. 2021. Semi-supervised semantic segmentation with cross pseudo supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2613–2622.

Chen, Y.; Zhu, X.; Li, W.; and Gong, S. 2020. Semi-supervised learning under class distribution mismatch. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 3569–3576.

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.

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.

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.

Ghosh, A.; and Thiery, A. H. 2020. On Data-Augmentation and Consistency-Based Semi-Supervised Learning. In *International Conference on Learning Representations*.

Guan, J.; Lai, R.; Xiong, A.; Liu, Z.; and Gu, L. 2020. Fixed pattern noise reduction for infrared images based on cascade residual attention CNN. *Neurocomputing*, 377: 301–313.

Han, J.; Ma, Y.; Zhou, B.; Fan, F.; Liang, K.; and Fang, Y. 2014. A robust infrared small target detection algorithm based on human visual system. *IEEE Geoscience and Remote Sensing Letters*, 11(12): 2168–2172.

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.

Hou, X.; and Zhang, L. 2007. Saliency detection: A spectral residual approach. In *2007 IEEE Conference on Computer Vision and Pattern Pecognition*, 1–8. IEEE.

Hu, H.; Wei, F.; Hu, H.; Ye, Q.; Cui, J.; and Wang, L. 2021. Semi-Supervised Semantic Segmentation via Adaptive Equalization Learning. In Ranzato, M.; Beygelzimer, A.; Dauphin, Y.; Liang, P.; and Vaughan, J. W., eds., *Advances in Neural Information Processing Systems*, volume 34, 22106–22118. Curran Associates, Inc.

Jin, Y.; Wang, J.; and Lin, D. 2022. Semi-Supervised Semantic Segmentation via Gentle Teaching Assistant. In Koyejo, S.; Mohamed, S.; Agarwal, A.; Belgrave, D.; Cho, K.; and Oh, A., eds., *Advances in Neural Information Processing Systems*, volume 35, 2803–2816. Curran Associates, Inc.

Kirchhoff, G. 1860. I. On the relation between the radiating and absorbing powers of different bodies for light and heat. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 20(130): 1–21.

Li, B.; Wang, L.; Wang, Y.; Wu, T.; Lin, Z.; Li, M.; An, W.; and Guo, Y. 2024. Mixed-Precision Network Quantization for Infrared Small Target Segmentation. *IEEE Transactions on Geoscience and Remote Sensing*.

Li, B.; Xiao, C.; Wang, L.; Wang, Y.; Lin, Z.; Li, M.; An, W.; and Guo, Y. 2023. Dense nested attention network for infrared small target detection. *IEEE Transactions on Image Processing*, 32: 1745–1758.

Li, J.; Socher, R.; and Hoi, S. C. 2019. DivideMix: Learning with Noisy Labels as Semi-supervised Learning. In *International Conference on Learning Representations*.

Li, S.; Xia, X.; Zhang, H.; Zhan, Y.; Ge, S.; and Liu, T. 2022. Estimating noise transition matrix with label correlations for noisy multi-label learning. *Advances in Neural Information Processing Systems*, 35: 24184–24198.

Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; and Dollár, P. 2017. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, 2980–2988.

Liu, Y.-C.; Ma, C.-Y.; He, Z.; Kuo, C.-W.; Chen, K.; Zhang, P.; Wu, B.; Kira, Z.; and Vajda, P. 2020. Unbiased Teacher for Semi-Supervised Object Detection. In *International Conference on Learning Representations*.

Na, J.; Ha, J.-W.; Chang, H. J.; Han, D.; and Hwang, W. 2023. Switching Temporary Teachers for Semi-Supervised Semantic Segmentation. In Oh, A.; Naumann, T.; Globerson, A.; Saenko, K.; Hardt, M.; and Levine, S., eds., *Advances in Neural Information Processing Systems*, volume 36, 40367–40380. Curran Associates, Inc.

Paul, H.; Greenberger, D. M.; Stenholm, S. T.; and Schleich, W. P. 2015. The Stefan–Boltzmann law: two classical laws give a quantum one. *Physica Scripta*, 2015(T165): 014027.

Redmon, J.; and Farhadi, A. 2018. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.

Ren, S.; He, K.; Girshick, R.; and Sun, J. 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, 234–241. Springer.

Sohn, K.; Berthelot, D.; Carlini, N.; Zhang, Z.; Zhang, H.; Raffel, C. A.; Cubuk, E. D.; Kurakin, A.; and Li, C.-L. 2020. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *Advances in neural information processing systems*, 33: 596–608.

Sun, R.; Mai, H.; Zhang, T.; and Wu, F. 2023. DAW: Exploring the Better Weighting Function for Semi-supervised Semantic Segmentation. In Oh, A.; Naumann, T.; Globerson, A.; Saenko, K.; Hardt, M.; and Levine, S., eds., *Advances in Neural Information Processing Systems*, volume 36, 61792–61805. Curran Associates, Inc.

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, Y.; Chen, H.; Heng, Q.; Hou, W.; Fan, Y.; Wu, Z.; Wang, J.; Savvides, M.; Shinozaki, T.; Raj, B.; et al. 2023. Freematch: Self-adaptive thresholding for semi-supervised learning. In *Proceedings of the International Conference On Learning Representations*, 7236–7246.

Xie, Z.; Sun, H.; and Li, M. 2023. Semi-supervised learning with support isolation by small-paced self-training. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 10510–10518.

Xu, H.; Liu, L.; Bian, Q.; and Yang, Z. 2022. Semi-supervised Semantic Segmentation with Prototype-based Consistency Regularization. In Koyejo, S.; Mohamed, S.; Agarwal, A.; Belgrave, D.; Cho, K.; and Oh, A., eds., *Advances in Neural Information Processing Systems*, volume 35, 26007–26020. Curran Associates, Inc.

Yang, L.; Qi, L.; Feng, L.; Zhang, W.; and Shi, Y. 2023. Revisiting weak-to-strong consistency in semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7236–7246.

Yang, L.; Zhuo, W.; Qi, L.; Shi, Y.; and Gao, Y. 2022. St++: Make self-training work better for semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4268–4277.

Zhang, B.; Wang, Y.; Hou, W.; Wu, H.; Wang, J.; Okumura, M.; and Shinozaki, T. 2021. Flexmatch: Boosting semi-supervised learning with curriculum pseudo labeling. *Advances in Neural Information Processing Systems*, 34: 18408–18419.

Zhang, L.; Peng, L.; Zhang, T.; Cao, S.; and Peng, Z. 2018. Infrared small target detection via non-convex rank approximation minimization joint l2, 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.; Wang, N.; Li, Y.; and Gao, X. 2019a. Deep latent low-rank representation for face sketch synthesis. *IEEE transactions on neural networks and learning systems*, 30(10): 3109–3123.

Zhang, M.; Wang, N.; Li, Y.; and Gao, X. 2019b. Neural probabilistic graphical model for face sketch synthesis. *IEEE transactions on neural networks and learning systems*, 31(7): 2623–2637.

Zhang, M.; Yang, H.; Guo, J.; Li, Y.; Gao, X.; and Zhang, J. 2024a. IRPruneDet: efficient infrared small target detection via wavelet structure-regularized soft channel pruning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 7224–7232.

Zhang, M.; Yue, K.; Li, B.; Guo, J.; Li, Y.; and Gao, X. 2024b. Single-frame Infrared Small Target Detection via Gaussian Curvature Inspired Network. *IEEE Transactions on Geoscience and Remote Sensing*.

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.

Zhang, M.; Zhang, R.; Zhang, J.; Guo, J.; Li, Y.; and Gao, X. 2023. Dim2Clear network for infrared small target detection. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 1–14.