EFLNet: Enhancing Feature Learning Network for Infrared Small Target Detection

Bo Yang, Xinyu Zhang, Jian Zhang, Jun Luo, Mingliang Zhou, *Member, IEEE*, and Yangjun Pi

***Abstract*— Single-frame infrared small target detection is considered to be a challenging task, due to the extreme imbalance between target and background, bounding box regression is extremely sensitive to infrared small target, and target information is easy to lose in the high-level semantic layer. In this article, we propose an enhancing feature learning network (EFLNet) to address these problems. First, we notice that there is an extremely imbalance between the target and the background in the infrared image, which makes the model pay more attention to the background features rather than target features. To address this problem, we propose a new adaptive threshold focal loss (ATFL) function that decouples the target and the background, and utilizes the adaptive mechanism to adjust the loss weight to force the model to allocate more attention to target features. Second, we introduce the normalized Gaussian Wasserstein distance (NWD) to alleviate the difficulty of convergence caused by the extreme sensitivity of the bounding box regression to infrared small target. Finally, we incorporate a dynamic head mechanism into the network to enable adaptive learning of the relative importance of each semantic layer. Experimental results demonstrate our method can achieve better performance in the detection performance of infrared small target compared to the state-of-the-art (SOTA) deep-learning-based methods. The source codes and bounding box annotated datasets are available at https://github.com/YangBo0411/infrared-small-target.**

***Index Terms*— Adaptive threshold focal loss (ATFL), deep learning, dynamic head, infrared small target detection.**

I. Introduction

INFRARED small target detection serves a crucial role in various applications, including ground monitoring [1], early warning systems [2], precision guidance [3], and others. In comparison to conventional object detection tasks, infrared small target detection exhibits distinct characteristics. First, due to the target's size or distance, the proportion of the target within the infrared image is exceedingly small, often comprising just a few pixels or a single pixel in extreme cases. Second, the objects in infrared small target detection tasks are typically sparsely distributed, usually containing only one or a few instances, each of which occupies a minuscule portion of the entire image. As a result, a significant imbalance arises between the target area and the background area. Moreover, the background of infrared small target is intricate, containing substantial amounts of noise and exhibiting a low signal-to-clutter ratio (SCR). Consequently, the target becomes prone to being overshadowed by the background. These distinctive features render infrared small target detection exceptionally challenging.

Various model-based methods have been proposed for infrared small target detection, including filter-based methods [4], [5], local contrast-based methods [6], [7], and low-rank-based methods [8], [9]. The filter-based methods segment the target by estimating the background and enhancing the target. However, their suitability is limited to even backgrounds, and they lack robustness when faced with complex backgrounds. The local contrast-based methods identify the target by calculating the intensity difference between the target and its surrounding neighborhood. Nevertheless, they struggle to effectively detect dim targets. The low-rank decomposition methods distinguish the structural features of the target and background based on the sparsity of the target and the low-rank characteristics of the background. Nonetheless, they exhibit a high false alarm (FA) rate when confronted with images featuring complex background and variations in target shape. In practical scenarios, infrared images often exhibit complex background, dim targets, and a low SCR, which poses a possibility of failure for these methods.

In recent years, deep learning has witnessed remarkable advancements, leading to significant breakthroughs in numerous domains. In contrast to traditional methods for infrared small target detection, deep learning leverages a data-driven end-to-end learning framework, enabling adaptive feature learning of infrared small target without the need for manual feature making. Since the work of miss detection versus FA (MDvsFA) [10] and asymmetric contextual modulation networks (ACMNets) [11], some deep-learning-based methods have been proposed. Despite the notable achievements of existing deep-learning-based methods in infrared small target detection, the majority of current research treats it as a segmentation task [12], [13], [14], [15]. The segmentation tasks offer pixel-level detailed information, which is advantageous in scenarios that demand precise differentiation. However, segmentation tasks necessitate the processing of pixel-level details, requiring substantial computational resources. Consequently, the training and inference times tend to be prolonged.

Manuscript received 31 October 2023; revised 14 December 2023 and 14 January 2024; accepted 8 February 2024. Date of publication 19 February 2024; date of current version 23 February 2024. This work was supported in part by the Fundamental Research Funds for the Central Universities under Grant 2023CDJXY-021. *(Corresponding author: Yangjun Pi.)*

Bo Yang, Xinyu Zhang, Jian Zhang, Jun Luo, and Yangjun Pi are with the State Key Laboratory of Mechanical Transmission for Advanced Equipment, College of Mechanical and Vehicle Engineering, Chongqing University, Chongqing 400044, China (e-mail: cqpp@cqu.edu.cn).

Mingliang Zhou is with the School of Computer Science, Chongqing University, Chongqing 400044, China.

Digital Object Identifier 10.1109/TGRS.2024.3365677

In addition, semantic segmentation is only an intermediate representation, which is used as input to track and locate infrared small target, and segmentation integrity is only an approximation of detection accuracy, and the specific detection performance cannot be evaluated. Therefore, there have been works to model infrared small target detection as an object detection problem [16], [17], [18], [19].

However, the detection performance of infrared small target remains insufficient compared to the detection of normal targets. This inadequacy can be attributed to three key factors. First, the imbalance between the target and the background in the image will cause the detector to learn more background information and tend to mistakenly recognize the target as the background, while not paying enough attention to the target information. Second, infrared small target are highly sensitive to the intersection over union (IoU) metric, rendering precise bounding box regression challenging as even slight changes in the bounding box can significantly impact the IoU calculation. Third, the information of infrared small target is easily lost during the downsampling process, and shallow features containing more target information are not taken seriously.

To address above problem, this article proposes a detection-based method called enhancing feature learning network (EFLNet), which can improve the detection performance of infrared small target. First, we design the adaptive threshold focal loss (ATFL) function to alleviate the imbalance problem between the target and the background in the infrared image. Furthermore, to achieve more accurate bounding box regression for infrared small target, a two-dimensional Gaussian distribution is used to remodel the bounding box, and the normalized Gaussian Wasserstein distance (NWD) is employed to address the problem of infrared small target being highly sensitive to IoU. Finally, we incorporate a dynamic head into the detection network. The relative importance of each semantic layer is learned through the self-attention mechanism, which improves the detection performance of infrared small target. Furthermore, most of the existing infrared small target datasets solely offer mask annotation versions, limiting the scope of infrared small target detection to a segmentation task. We provide the corresponding bounding box annotation versions for the current infrared small target public datasets, which makes it possible to make infrared small target detection as a detection-based task.

Our contributions can be summarized as follows.

1) We propose an EFLNet to improve the detection performance of infrared small target. The feature learning ability of the network for infrared small target can be well enhanced by more suitable loss function and network structure.
2) We designed an ATFL for infrared small target, which can decouple the target from the background and dynamically adjust the loss weight, allowing the model to assign greater attention to hard-to-detect targets.
3) We provide a bounding box annotation version of the current infrared small target public dataset, which makes up for the lack of bounding box annotation version in the current dataset and facilitates the detection task.

The remainder of this article is organized as follows. Section II provides related work of the existing research on infrared small target detection. Section III introduces the proposed network architecture. The experimental results and analysis are presented in Section IV. Finally, Section V concludes the entire article.

II. Related Work

A. Model-Based Method

Extensive research was conducted by researchers to address the problem of infrared small target detection. Filter-based methods, such as MaxMedian [4], Tophat [5], 2-D adaptive least-mean-square (TDLMS) [20], and 2-D variational mode decomposition (TDVMD) [21], demonstrated good performance on smooth or low-frequency backgrounds but exhibited limitations when dealed with complex backgrounds. Local-contrast-based methods like weighted strengthened local contrast measure (WSLCM) [2], trilayer local contrast measure (TLLCM) [22], improved local contrast measure (ILCM) [6], and relative local contrast measure (RLCM) [7] assumed that the target's brightness was higher than its neighborhood, thereby failed to effectively detect dim targets. On the other hand, low-rank decomposition-based methods, including infrared patch-image (IPI) [8], nonconvex rank approximation minimization joint $l_{2,1}$ norm (NRAM) [9], reweighted infrared patch-tensor (RIPT) [23], and partial sum of the tensor nuclear norm (PSTNN) [24], achieved target background separation based on the assumption of a low-rank background and sparse target, but they were susceptible to background clutter and lack strong adaptability. However, real-world scenes often exhibit a high level of background complexity, characterized by clutter and noise. Moreover, the target typically manifests as a faint feature due to the long imaging distance. Consequently, the performance of conventional methods is hindered by these limitations, leading to poor detection performance in real-world scenarios.

B. Deep-Learning-Based Method

Data-driven methods leveraging deep learning techniques demonstrated the ability to adaptively extract features from images and acquire high-level semantic information. Accordingly, the deep-learning-based methods exhibited superior performance compared to traditional approaches when confronted with various complex environments. Moreover, with the opening of numerous infrared small target datasets attracted increasing interest among researchers on deep-learning-based methods. Based on distinct processing paradigms, deep-learning-based approaches can be categorized into two main groups: detection-based and segmentation-based methods.

1) Segmentation-Based Methods: Segmentation-based methods employed pixel-by-pixel threshold segmentation on the image, yielded a segmentation mask that provided object position and size information. Wang et al. [10] introduced a generative adversarial network (GAN) framework for adversarial learning, enabling the natural attainment of Nash equilibrium between miss detection (MD) and FA during

training. Dai et al. [11] proposed an asymmetric contextual modulation (ACM) module that combined top-down and bottom-up point-wise attention mechanisms to enhance the encoding of semantic information and spatial details. Additionally, Dai et al. [25] presented a model-driven deep network attentional local contrast networks (ALCNets) that effectively utilized labeled data and domain knowledge, addressed issues such as inaccurate modeling, hyperparameter sensitivity, and insufficient intrinsic features. Zhang et al. [26] introduced the infrared shape network (ISNet) for detecting shape information in infrared small target. To mitigate deep information loss caused by pooling layers in infrared small target, Li et al. [27] proposed the dense nested attention network (DNA-Net). Hou et al. [28] devised a robust infrared small target detection network (RISTDNet) that combined handcrafted feature methods with CNN. Chen et al. [29] developed a hierarchical overlapped small patch transformer (HOSPT) as a replacement for convolution kernels in convolutional neural network (CNN), enabled the encoding of multiscale features and addressed the challenge of modeling long-range dependencies in images.

2) Detection-Based Methods: Detection-based methods were the same as the ordinary object detection algorithms, they directly outputed the target's position and scale information. To enhance the detection performance of infrared small target, Li and Shen [30] proposed a method that incorporated super-resolution enhancement of the input image and improved the structure of YOLOv5. In a similar vein, Zhou et al. [31] tackled the challenge of detecting infrared small target by employing a YOLO-based framework. Dai et al. [16] introduced a one-stage cascade refinement network (OSCAR) to address the issues of inherent characteristics deficiency and inaccurate bounding box regression in infrared small target detection. Meanwhile, Yao et al. [32] developed a lightweight network that combined traditional filtering methods with the standard convolutional one stage object detection (FCOS) to improve responsiveness to infrared small target. Du et al. [17] adopted an interframe energy accumulation (IFEA) enhancement mechanism to amplify the energy of moving time series targets. Furthermore, the issue of sample misidentification was resolved by employing a small IoU strategy. Similarly, Ju et al. [33] achieved the same objective through the utilization of an image filtering module.

III. Methodology

A. Overall Architecture

Fig. 1 shows the workflow of the proposed method. First, the infrared image serves as the input to the backbone network, enabling the extraction of essential features. These features undergo fusion via FPN and PAN [34], integrating multi-scale information. The resulting fused features are then fed into the dynamic detection head, facilitating the learning of the relative significance of diverse semantic layers. Ultimately, the detection results are assessed by the NWD and ATFL, which compute the loss and guide the model optimization process.

B. Adaptive Threshold Focal Loss

The infrared image predominantly consists of background, with only a small portion occupied by the target, as illustrated in Fig. 2. Thus, learning the characteristics of the background during the training process is easier than learning those of the target. The background can be considered as easy samples, while the targets can be regarded as hard samples. However, even the well-learned background still produces losses during training. In fact, the background samples that occupy the main part of the infrared image dominate the gradient update direction, overwhelmed the target information. To address this issue, we propose a new ATFL function. First, the threshold setting is used to decouple the easy-to-identify background from the difficult-to-identify target. Second, by intensifying the loss associated with the target and mitigating the loss linked to the background, we force the model to allocate greater attention to target features, thereby alleviating the imbalance between the target and the background. Finally, adaptive design has been applied to hyperparameters to reduce time consumption caused by adjusting hyperparameters.

We propose an ATFL that decouples the target and background based on the set threshold. The loss value is adaptively adjusted according to the predicted probability value, aiming to enhance the detection performance of infrared small target.

The classical cross-entropy loss function can be expressed as

$$\mathcal{L}_{\mathrm{BCE}} = -(y\log(p) + (1-y)\log(1-p)) \tag{1}$$

where p represents the predicted probability and y represents the true label. Its succinct representation is

$$\mathcal{L}_{\mathrm{BCE}} = -\log(p_t) \tag{2}$$

where

$$p_t = \begin{cases} p, & \text{if } \ y = 1 \\ 1-p, & \text{others.} \end{cases} \tag{3}$$

The cross-entropy function cannot address the imbalance problem between samples, so the focal loss [35] function introduces a modulation factor $(1-p_t)^{\gamma}$ to reduce the loss contribution of easily classifiable samples by adjusting the focusing parameter γ. The focus loss function can be expressed as

$$\mathrm{FL}(p_{\mathrm{t}}) = -(1-p_{\mathrm{t}})^{\gamma}\log(p_{\mathrm{t}}). \tag{4}$$

The focal loss function can adjust the value of the γ to reduce the loss weight of easy samples, as can be seen in Fig. 3. However, while reducing the loss of easy samples, the modulation factor also reduces the value of difficult sample losses, which is not conducive to the learning of difficult samples.

To address above problem, we propose a threshold focal loss (TFL) function, which effectively mitigates the impact of easy samples by reducing their loss weight, while simultaneously increasing the loss weight assigned to difficult samples. Specifically, we designate prediction probability value above 0.5 as

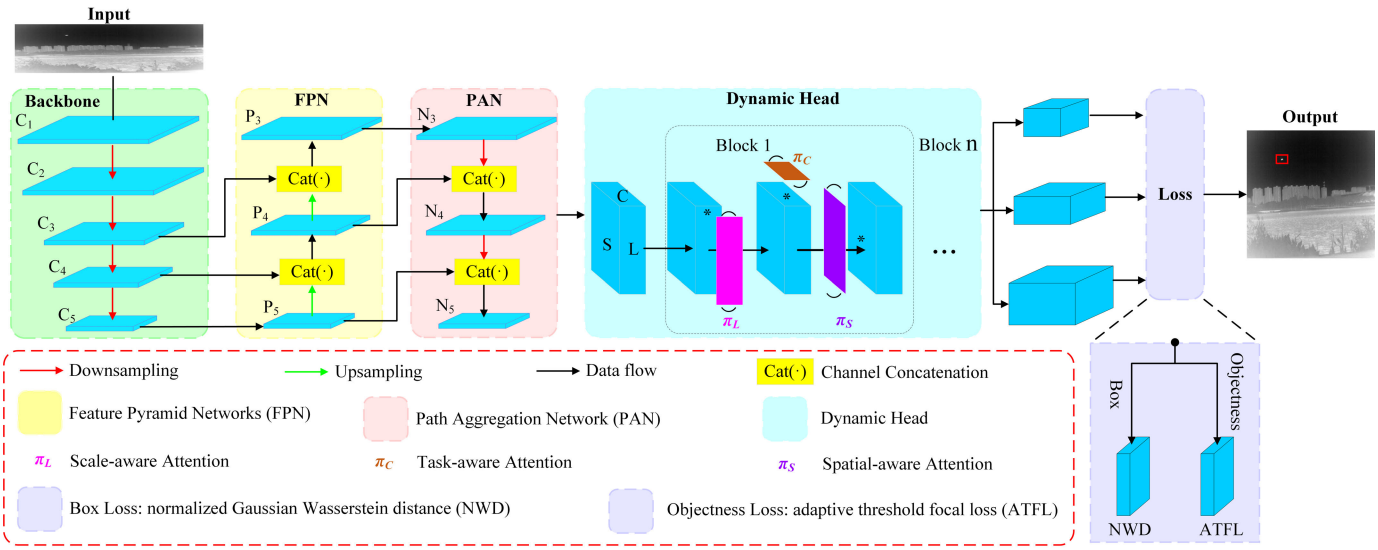

Fig. 1. Overview of the proposed EFLNet, which has the structure of backbone, FPN, PAN, and dynamic head, as well as the loss functions of NWD and ATFL.

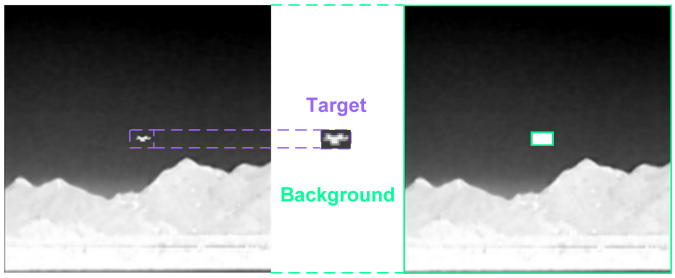

Fig. 2. Imbalance phenomenon between the target and the background.

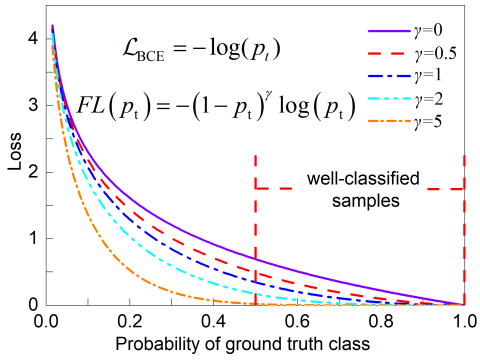

Fig. 3. Changes in losses in terms of different γ. The $p_t > 0.5$ is regarded as well-classified samples.

easy samples, while conversely considering values below this threshold as hard samples. The expression is as follows:

$$\mathrm{TFL} = \begin{cases} -(\lambda - p_t)^{\eta}\log(p_t), & p_t <= 0.5 \\ -(1-p_t)^{\gamma}\log(p_t), & p_t > 0.5 \end{cases} \tag{5}$$

where η, γ, and $\lambda(>1)$ are the hyperparameters. For different datasets and models, the hyperparameters need to be adjusted multiple times to achieve optimal performance. In the field of artificial intelligence, each training takes a lot of time, resulting in expensive time costs. Therefore, we have made adaptive improvements to η and γ.

For easy samples, we expect the loss value to decrease as p_t increases, further reducing the loss generated by easy samples. At the beginning of training, even easy samples will have a relatively low prediction probability and gradually rise as the training process progresses, and γ should gradually approach 0. The predicted probability value $\hat{p}_c$ of the real target can be used to mathematically model the progress of model training, and it can be predicted by exponential smoothing. It is stated as follows:

$$\hat{p}_c = 0.05 \times \frac{1}{t-1}\sum_{i=0}^{t-1}\overline{p_i} + 0.95 \times p_t \tag{6}$$

where $\hat{p}_c$ represents the predicted value for the next epoch, p_t represents the current average predicted probability value, and $\overline{p}_i$ represents the average predicted probability value for each training epoch. According to Shannon's information theory, the greater the probability value of an event, the smaller the amount of information it brings; Conversely, the greater the amount of information. Thus, the adaptive modulation factor γ can be expressed as

$$\gamma = -\ln(\hat{p}_c). \tag{7}$$

However, in the later stage of network training, the expected probability value is too large, which will reduce the proportion of difficult samples. We express the η as

$$\eta = -\ln(p_t). \tag{8}$$

By incorporating (7), (8) into (5), the expression of the ATFL can be obtained as

$$\mathrm{ATFL} = \begin{cases} -(\lambda - p_t)^{-\ln(p_t)}\log(p_t), & p_t <= 0.5 \\ -(1-p_t)^{-\ln(\hat{p}_c)}\log(p_t), & p_t > 0.5. \end{cases} \tag{9}$$

C. Normalized Gaussian Wasserstein Distance

The IoU metric used for ordinary object detection exhibits extreme sensitivity when applied to infrared small target. Even a slight deviation in position between the predicted boxes

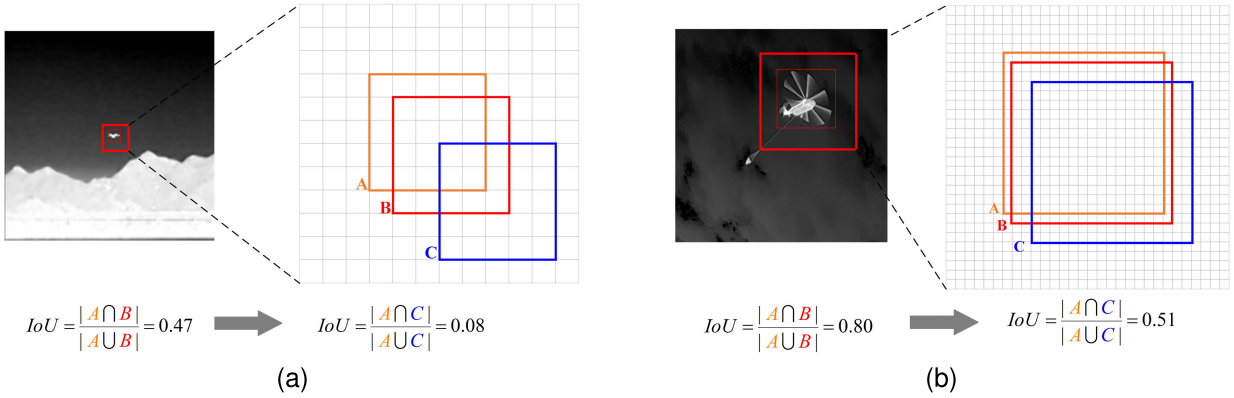

Fig. 4. Sensitivity analysis of IoU on tiny and normal scale object. (a) Tiny scale object. (b) Normal scale object.

and ground-truth boxes can result in a significant change in IoU. This sensitivity is illustrated in Fig. 4, where a small position deviation leads to a decrease in IoU for infrared small target from 0.47 to 0.08. Conversely, for normal-sized objects, the IoU only decreases from 0.80 to 0.51 under the same position deviation. Such sensitivity of the IoU metric toward infrared small target leads to a high degree of similarity between positive and negative samples during training, making it challenging for the network to converge effectively. Furthermore, in extreme cases, the infrared small target may occupy only one or a few pixels within the image. Consequently, the IoU between the ground truth and any predicted bounding box falls below the minimum threshold, resulting in zero-positive samples within the image. Therefore, alternative evaluation indicators are required for assessing infrared small target more accurately.

The IoU metric is actually a similarity calculation between samples, which is sensitive to the size change of the target and is not suitable for infrared small target, so we introduce NWD as a new measure. The Wasserstein distance can measure the similarity between distributions with minimal or no overlap, and it is also insensitive to objects of different scales. Therefore, it can address issues related to the similarity of positive and negative samples, as well as sparse positive samples during the training process of infrared small target. Specifically, the bounding box is modeled as a 2-D Gaussian distribution

$$f\left(\mathbf{x}\middle|\boldsymbol{\mu}, \boldsymbol{\Sigma}\right) = \frac{\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\top}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)}{2\pi\left|\Sigma\right|^{\frac{1}{2}}} \tag{10}$$

where $\mathbf{x}$, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ represent the coordinates (x, y), the mean vector, and covariance matrix of the Gaussian distribution. When

$$(\mathbf{x}-\boldsymbol{\mu})^{\top}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) = 1. \tag{11}$$

The horizontal bounding box $R = (c_x, c_y, w, h)$ can be modeled as a 2-D Gaussian distribution using $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$$\mu = \begin{bmatrix} c_x \\ c_y \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \frac{w^2}{4} & 0 \\ 0 & \frac{h^2}{4} \end{bmatrix} \tag{12}$$

where (c_x, c_y), w, and h represent the center coordinates, width, and height, respectively. The 2-D Wasserstein distance between two 2-D Gaussian distributions $\mu_1 = N(\boldsymbol{m_1}, \boldsymbol{\Sigma}_1)$ amd $\mu_2 = N(\boldsymbol{m_2}, \boldsymbol{\Sigma}_2)$ is defined as

$$W_2^2(\mu_1, \mu_2) = \left\|\mathbf{m}_1 - \mathbf{m}_2\right\|_2^2 + \mathbf{Tr}\left(\Sigma_1 + \boldsymbol{\Sigma}_2 - 2\left(\boldsymbol{\Sigma}_2^{1/2}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2^{1/2}\right)^{1/2}\right). \tag{13}$$

It can be simplified as

$$W_2^2(\mu_1, \mu_2) = \left\|\mathbf{m}_1 - \mathbf{m}_2\right\|_2^2 + \left\|\boldsymbol{\Sigma}_1^{1/2} - \boldsymbol{\Sigma}_2^{1/2}\right\|_F^2 \tag{14}$$

where $\|\cdot\|_F$ is the Frobenius norm. The distance between the Gaussian distributions N_a, N_b modeled by bounding boxes $A = (cx_a, cy_a, w_a, h_a)$ and $B = (cx_b, cy_b, w_b, h_b)$ can be simplified as

$$W_2^2(\mathcal{N}_a, \mathcal{N}_b) = \left(\left[cx_a, cy_a, \frac{w_a}{2}, \frac{h_a}{2}\right]^{\mathrm{T}}, \left[cx_b, cy_b, \frac{w_b}{2}, \frac{h_b}{2}\right]^{\mathrm{T}}\right)_2^2. \tag{15}$$

Normalizing it exponentially to a range of 0–1 gives the normalized Watherstein distance [36]

$$\mathrm{NWD}(\mathcal{N}_a, \mathcal{N}_b) = \exp\left(-\frac{\sqrt{W_2^2(\mathcal{N}_a, \mathcal{N}_b)}}{C}\right) \tag{16}$$

where C is a constant related to the dataset.

D. Dynamic Head

Feature pyramid networks, which involve combining multi-scale convolution features, have become a prevalent technique in detection networks. Nevertheless, it is important to note that features at varying depths convey distinct semantic information. Specifically, during the down-sampling process, infrared small target may experience information loss. Shallow features contain valuable infrared small target information that merits greater attention from the network. On the other hand, different viewpoints and task forms will produce different features and target constraints, which bring difficulties to infrared small target detection. The dynamic head as shown in Fig. 5 can adaptively focus on the scale-space-task information of objects, which can better learn the relative importance of each semantic level and spatial information of the target, as well as adaptively match different task forms.

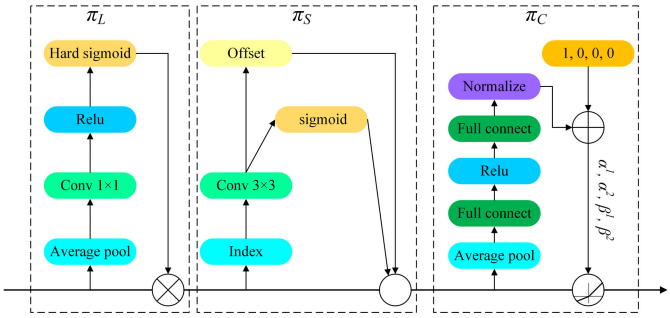

Fig. 5. Structure of dynamic head block. The π_L denotes scale-aware attention, π_S is spatial-aware attention, and π_C represents task-aware attention.

Given the feature tensor $F \in R^{L\times S\times C}$, L represents the number of pyramid layers, S represents the size of the feature, $S = H \times W$, H, W represents the height and width of the feature, and C represents the number of channels. Dynamic head [37] can be expressed as

$$W(\mathcal{F}) = \pi_C(\pi_S(\pi_L(\mathcal{F}) \cdot \mathcal{F}) \cdot \mathcal{F}) \cdot \mathcal{F} \tag{17}$$

where $\pi_L(\cdot)$, $\pi_S(\cdot)$, and $\pi_C(\cdot)$ represents the attention function on L, S, and C, respectively. Scale-aware attention π_L enables dynamic feature fusion based on the importance of features in each layer

$$\pi_L(\mathcal{F}) \cdot \mathcal{F} = \sigma\left(f\left(\frac{1}{SC}\sum_{S,C}\mathcal{F}\right)\right) \cdot \mathcal{F} \tag{18}$$

where $f(\cdot)$ is a 1×1 convolutional layer $\sigma(x) = \max(0, \min(1, (x + 1/2)))$ is a hard-sigmoid function.

Spatial-aware attention $\pi_S(\cdot)$ uses deformable convolution [38] to fuse features of different levels in the same spatial position

$$\pi_S(\mathcal{F}) \cdot \mathcal{F} = \frac{1}{L}\sum_{l=1}^{L}\sum_{k=1}^{K} w_{l,k} \cdot \mathcal{F}(l; p_k + \Delta p_k; c) \cdot \Delta m_k \tag{19}$$

where K is the number of sparse sampling locations, $p_k + \Delta p_k$ and Δm_k are learned from the input features, $p_k + \Delta p_k$ is a shifted location by the self-learned spatial offset Δp_k, Δm_k is an important scalar for self-learning at position p_k. Task-aware attention $\pi_C(\cdot)$ dynamically switches ON and OFF channels to support different tasks

$$\pi_C(\mathcal{F}) \cdot \mathcal{F} = \max\left(\alpha^1(\mathcal{F}) \cdot \mathcal{F}_c + \beta^1(\mathcal{F}), \alpha^2(\mathcal{F}) \cdot \mathcal{F}_c + \beta^2(\mathcal{F})\right) \tag{20}$$

where $[\alpha^1, \alpha^2, \beta^1, \beta^2]^T$ is a superfunction that controls the threshold, which reduces the dimension in the $L \times S$ dimension through average pooling, then uses two fully connected layers and a normalization layer, and finally normalizes by the sigmoid activation function.

IV. Experiment

A. Dataset

1) Datasets: We conducted experiments using bounding box annotation and semantic segmentation mask annotation from three publicly available infrared small target datasets (as depicted in Fig. 6): NUAA-SIRST [11], NUDT-SIRST [27], and IRSTD-1k [26]. To ensure proper evaluation, we divided each dataset into training set, validation set and test set, following a ratio of 6:2:2.

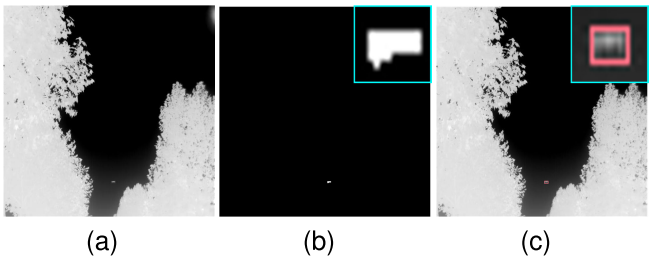

Fig. 6. Different annotation forms for the current infrared small target public dataset. (a) Image. (b) Semantic segmentation. (c) Bounding box.

2) Evaluation Metrics: To compare the proposed method with the state-of-the-art (SOTA) methods, we employ commonly used evaluation metrics including precision, recall, and F1. Each metric is defined as follows:

Precision: Precision is calculated as the ratio of true positives (TPs) to the sum of TPs and false positives (FPs)

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP}). \tag{21}$$

Recall: Recall is calculated as the ratio of TPs to the sum of TPs and false negatives (FNs)

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN}). \tag{22}$$

$F1$: F1 is a harmonic mean of precision and recall, providing a balanced measure of the model's performance, which is computed as

$$F1 = 2 \times (\text{Precision} * \text{Recall})/(\text{Precision} + \text{Recall}). \tag{23}$$

3) Comparison to the SOTA Methods: Deep-learning methods have exhibited significantly superior performance compared to model-based methods, such as Top-Hat, Max-Median, WSLCM, TLLCM, NRAM, IPI, and RIPT. These results have been widely demonstrated [10], [25], [26], [27]. Therefore, this article does not compare with model-based methods anymore, but with several deep-learning-based most advanced methods, including MDvsFA [10], AGPCNet [39], ACM [11], ISNet [26], ALCNet [25], DNANet [27]. To ensure fair comparisons, each model has been retrained using a repartitioned dataset, employing a training epoch of 400 and keeping the remaining parameters at their default values.

B. Quantitative Results

Table I presents a quantitative comparison of the results obtained from different methods. The proposed method demonstrates the highest performance across all evaluation metrics on the NUAA-SIRST, NUDT-SIRST, and IRSTD-1k datasets when compared to the SOTA method which proves the effectiveness of the proposed method. Because most of the current deep-learning-based methods regard infrared small target detection as a pixel-level segmentation task, pixel-level segmentation results need to be obtained. The slightest inadequacy will produce FAs or missed detections, resulting in poor

TABLE I

COMPARISONS WITH SOTA METHODS ON NUAA-SIRST, NUDT-SIRST, AND IRSTD-1K IN PRECISION, RECALL, AND F1

Method	NUAA-SIRST			NUDT-SIRST			IRSTD-1k		
	Precision	*Recall*	$F1$	*Precision*	*Recall*	$F1$	*Precision*	*Recall*	$F1$
MDvsFA[9]	0.845	0.507	0.597	0.608	0.192	0.262	0.550	0.483	0.475
AGPCNet[30]	0.390	0.810	0.527	0.368	0.684	0.479	0.415	0.470	0.441
ACM[10]	0.765	0.762	0.763	0.732	0.745	0.738	0.679	0.605	0.640
ISNet[18]	0.820	0.847	0.834	0.742	0.834	0.785	0.718	0.741	0.729
ACLNet[17]	0.848	0.78	0.813	0.868	0.772	0.817	0.843	0.656	0.738
DNANet[19]	0.847	0.836	0.841	0.914	0.889	0.901	0.768	0.721	0.744
Ours	**0.882**	**0.858**	**0.870**	**0.963**	**0.931**	**0.947**	**0.870**	**0.817**	**0.843**

TABLE II

ABLATION STUDY ON THE DIFFERENT HYPERPARAMETER FORM OF ATFL IN PRECISION, RECALL, $AP_{0.5}$, AND F1

Hyperparameter form	η	γ	*Precision*	*Recall*	$AP_{0.5}$	$F1$
Fixed hyperparameters	2	2	0.875	0.723	0.762	0.792
	2	4	0.777	0.760	0.718	0.768
	2	6	0.780	0.762	0.728	0.771
	2	8	0.742	0.804	0.735	0.772
	2	10	0.736	0.727	0.702	0.731
	2	2	0.875	0.723	0.762	0.792
	4	2	0.889	0.698	0.767	0.782
	6	2	0.813	0.712	0.739	0.759
	8	2	0.816	0.669	0.715	0.735
	10	2	0.679	0.756	0.722	0.715
Adaptive hyperparameters	/	/	0.876	0.749	0.780	0.808

TABLE III

ABLATION STUDY ON THE DIFFERENT PARAMETER λ OF ATFL IN PRECISION, RECALL, $AP_{0.5}$, AND F1

λ	*Precision*	*Recall*	$AP_{0.5}$	$F1$
Baseline	0.845	0.756	0.770	0.798
1.5	0.851	0.768	0.773	0.807
2	0.879	0.726	0.763	0.795
2.5	0.876	0.749	0.780	0.808
3	0.889	0.745	0.763	0.810
3.5	0.851	0.790	0.790	0.819
4	0.870	0.736	0.748	0.797

TABLE IV

ABLATION STUDY ON THE DIFFERENT THRESHOLD SETTING OF ATFL IN PRECISION, RECALL, $AP_{0.5}$, AND F1

Threshold setting	*Precision*	*Recall*	$AP_{0.5}$	$F1$
Baseline	0.845	0.756	0.770	0.798
0.1	0.869	0.723	0.766	0.789
0.3	0.816	0.772	0.782	0.793
0.5	0.885	0.746	0.784	0.810
0.7	0.895	0.736	0.778	0.808
0.9	0.855	0.720	0.757	0.782

TABLE V

ABLATION STUDY ON THE DIFFERENT PARAMETER C OF NWD IN PRECISION, RECALL, $AP_{0.5}$, AND F1

C	*Precision*	*Recall*	$AP_{0.5}$	$F1$
Baseline	0.845	0.756	0.770	0.798
9	0.867	0.775	0.789	0.818
11	0.890	0.781	0.806	0.832
13	0.867	0.797	0.800	0.831
15	0.864	0.778	0.795	0.819
17	0.899	0.771	0.801	0.830

TABLE VI

ABLATION STUDY ON THE DIFFERENT NUMBER OF DYNAMIC HEAD BLOCKS IN PRECISION, RECALL, $AP_{0.5}$, AND F1

Block	*Precision*	*Recall*	$AP_{0.5}$	$F1$
0	0.867	0.775	0.789	0.818
1	0.850	0.804	0.797	0.826
2	0.896	0.772	0.794	0.829
3	0.881	0.788	0.792	0.832
4	0.861	0.814	0.799	0.837
5	0.860	0.814	0.800	0.836

detection performance at the target level. In addition, there is less attention paid to the imbalance phenomenon and boundary box sensitivity issues in infrared small target. Therefore, these methods often yield poor performance in target-level detection tasks, resulting in relatively low precision, recall, and F1 scores. We designed ATFL to solve the imbalance between target and background by adaptive adjustment of loss weight, and make the model better learn the features of infrared small target through NWD metric and dynamic head. Therefore, it shows better detection performance for infrared small target.

C. Visual Results

The partial visualization results of different methods on the NUAA-SIRST, NUDT-SIRST, and IRSTD-1k datasets are shown in Fig. 7. The areas corresponding to correctly detected targets, FAs, and missed detections are highlighted by circles in red, orange, and purple, respectively. The red circle indicates the correct target, the orange circle indicates the FA, and the purple circle indicates the missed detection. A model with superior detection performance exhibits a greater number of red circles and fewer orange and purple circles in the graph. Generally, deep learning-based methods exhibit robust performance owing to their adaptive learning features, enabling effective detection of the majority of targets. However, most of these methods tend to produce FAs when encountering locally highlighted interference [as shown in Fig. 7 (1), (4), (5)]. Additionally, missed detection occurs when the target appears

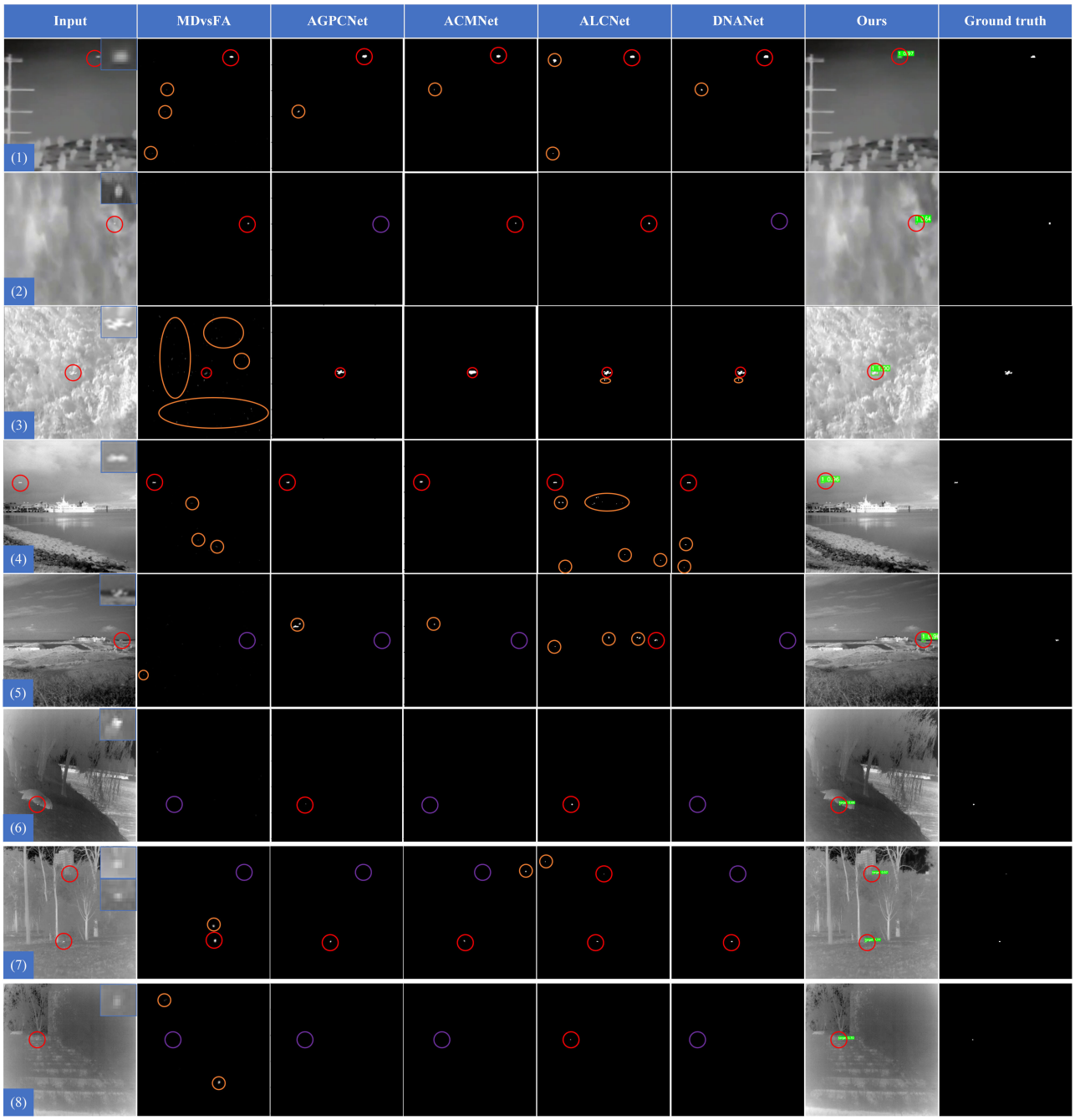

Fig. 7. Partial visual results gained by different methods on NUAA-SIRST, NUDT-SIRST, and IRSTD-1k datasets. The targets are represented by circles colored in red, purple, and orange, indicating correctly detected targets, miss detected targets, and false detected targets, respectively.

dim [as depicted in Fig. 7 (5), (7), (8)]. Our proposed method effectively learns the characteristics of infrared small target, allowing for accurate detection and localization even in the presence of local highlight interference and dim targets.

D. Ablation Study

We selected the IRSTD-1k dataset as our experimental dataset due to its composition of real images and an ample quantity of data. The NUAA dataset contains a limited number of images, while the NUDT dataset consists of images generated through simulation. To assess the effectiveness of each component within the EFLNet, we conducted multiple ablation experiments on the IRSTD-1k dataset.

1) Impact of ATFL: We investigated the effects of different hyperparameter forms and different values of λ on ATFL. As shown in Table II, when using a fixed hyperparameter form, multiple adjustments to η and γ are required, which can be time-consuming. On the contrary, when employing adaptive hyperparameter, the optimized results can be obtained with a single tuning operation, eliminating the need for multiple parameter adjustments. Furthermore, the adaptive mechanism yielded superior results compared to fixed hyperparameter. Thus, the effectiveness of our designed adaptive mechanism

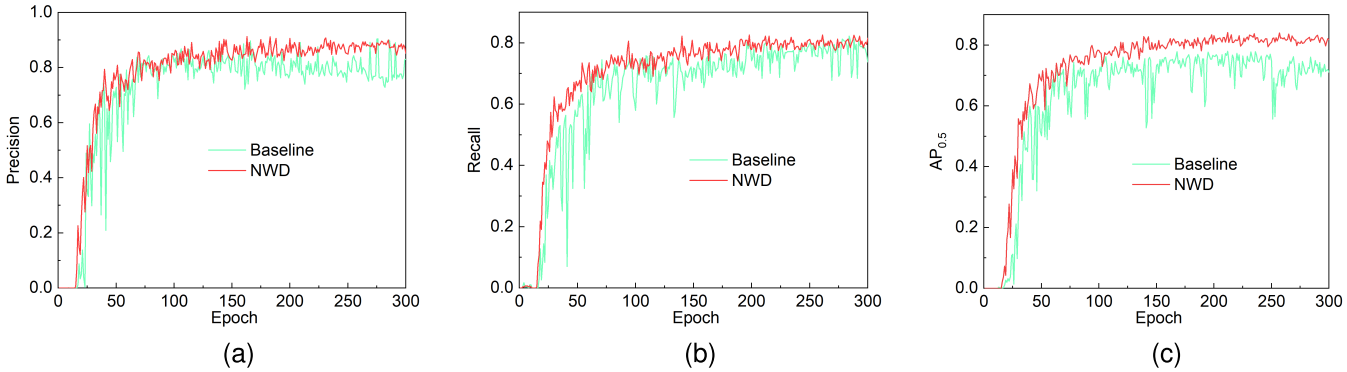

Fig. 8. Baseline and NWD comparison in terms of precision, recall, and $AP_{0.5}$ in training process. (a) Precision comparison. (b) Recall comparison. (c) $AP_{0.5}$ comparison.

TABLE VII
ABLATION STUDY ON THE ATFL, NWD, AND DYNAMIC HEAD IN PRECISION, RECALL, $AP_{0.5}$, F1, AND GFLOPS

Dataset	ATFL	NWD	Dynamic head	*Precision*	*Recall*	$AP_{0.5}$	F1	Parameters(M)
IRSTD	×	×	×	0.845	0.756	0.770	0.798	32.769
	✓	×	×	0.851	0.768	0.773	0.807	32.769
	✓	✓	×	0.858	0.797	0.801	0.826	32.769
	✓	✓	✓	**0.882**	**0.807**	**0.816**	**0.843**	38.519

has been validated. As shown in Table III, we change the λ values (e.g., 1.5, 2, 2.5, 3, 3.5, 4) and compared their impact on model performance against the baseline. The initial baseline model exhibits a relatively low detection rate for targets (recall $=$ 0.743). However, with the incorporation of the ATFL, the performance of model undergoes a significant enhancement. By assigning greater importance to hard-to-detect targets, the detection rate of infrared small target is improved, resulting in an enhanced recall rate of up to 0.790. Notably, when $\lambda = 3.5$, the overall performance reaches its optimal level, validating the effectiveness of our method. As can be seen from Table IV that overly small or large thresholds result in decreased performance. This can be attributed to the imprecise classification of samples when the threshold is set unreasonable, potentially leading to negative effect. The optimal performance is achieved when the threshold is set at 0.5.

2) Impact of NWD: As mentioned previously, the parameter C is closely tied to the dataset. To investigate its influence on the model, we conducted experiments by varying the value of parameter C as shown in Table V. The NWD enhances the quality of both positive and negative samples during the training process. As a result, the application of NWD yields a significant improvement in the model's performance, and reaching its optimum at $C = 11$. Fig. 8 illustrates the changes in evaluation metrics during the training process. Since the IoU metric is particularly sensitive to infrared small target, leading to similarities between positive and negative samples. Accordingly, the model encounters difficulties in convergence, resulting in substantial fluctuations in the evaluation metric. However, as can be seen from the figure, the integration of NWD alleviates the problem of difficult in model convergence.

3) Impact of Dynamic Head: It can be seen that IoU metric can easily lead to the model not convergence, and it is difficult to evaluate the actual effect of the network. Therefore, we conducted experiments on dynamic head under the premise of using NWD. Table VI shows the obvious improvement in model performance after incorporating the dynamic head module. Moreover, as the quantity of dynamic head is augmented, the performance of the model will increase slightly. The optimal performance result can be achieved when the number of dynamic head module is 4.

In addition to analyzing the effectiveness of each component, we also experimented with the combined effects of multiple components. It can be seen from Table VII that only the dyhead raises the number of network parameters, and the design of the loss function does not additional increase the complexity of the model. Moreover, the performance can be improved with the addition of each component, indicating that the designed loss function and the network structure can be well integrated.

V. CONCLUSION

This article presented the EFLNet, an innovative approach aimed at enhancing the feature learning capability of infrared small target, thereby improving the performance of infrared small target detection. Specifically, we designed a novel ATFL loss function that automatically adjusted the loss weights, allowing for differentiated treatment of the target and background, which alleviated the inherent imbalance problem between the target and the background within the image. The NWD metric facilitated the generation of superior quality positive and negative samples, effectively resolving the sensitivity issues associated with the IoU metric when dealed with infrared small target. By leveraging dynamic head, the relative importance of each semantic layer can be learned, and more attention was paid to the shallow features of infrared small target. Experiments on public datasets showed that our method outperforms SOTA methods. Additionally, we provided the additional bounding box annotation forms of the existing infrared small target datasets, which makes it possible to treat infrared small target detection as a detection-based task.

REFERENCES

[1] K. Wang, S. Du, C. Liu, and Z. Cao, "Interior attention-aware network for infrared small target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5002013.

[2] J. Han et al., "Infrared small target detection based on the weighted strengthened local contrast measure," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 9, pp. 1670–1674, Sep. 2021.

[3] Y. Sun, J. Yang, and W. An, "Infrared dim and small target detection via multiple subspace learning and spatial–temporal patch-tensor model," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 3737–3752, May 2021.

[4] S. D. Deshpande, E. H. Meng, R. Venkateswarlu, and P. Chan, "Max-mean and max-median filters for detection of small targets," in *Signal and Data Processing of Small Targets*, vol. 3809. Bellingham, WA, USA: SPIE, 1999, pp. 74–83.

[5] J. F. Rivest and R. Fortin, "Detection of dim targets in digital infrared imagery by morphological image processing," *Opt. Eng.*, vol. 35, no. 7, pp. 1886–1893, Jul. 1996.

[6] J. Han, Y. Ma, B. Zhou, F. Fan, K. Liang, and Y. Fang, "A robust infrared small target detection algorithm based on human visual system," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 12, pp. 2168–2172, Dec. 2014.

[7] J. Han, K. Liang, B. Zhou, X. Zhu, J. Zhao, and L. Zhao, "Infrared small target detection utilizing the multiscale relative local contrast measure," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 4, pp. 612–616, Apr. 2018.

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

[9] L. Zhang, L. Peng, T. Zhang, S. Cao, and Z. Peng, "Infrared small target detection via non-convex rank approximation minimization joint $l_{2,1}$ norm," *Remote Sens.*, vol. 10, no. 11, p. 1821, 2018.

[10] H. Wang, L. Zhou, and L. Wang, "Miss detection vs. false alarm: Adversarial learning for small object segmentation in infrared images," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 8509–8518.

[11] Y. Dai, Y. Wu, F. Zhou, and K. Barnard, "Asymmetric contextual modulation for infrared small target detection," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, Oct. 2021, pp. 950–959.

[12] X. Ying et al., "Mapping degeneration meets label evolution: Learning infrared small target detection with single point supervision," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 15528–15538.

[13] B. Nian, B. Jiang, H. Shi, and Y. Zhang, "Local contrast attention guide network for detecting infrared small targets," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5607513.

[14] R. Li et al., "Direction-coded temporal U-shape module for multiframe infrared small target detection," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, pp. 1–14, 2023, doi: 10.1109/TNNLS.2023.3331004.

[15] H. Sun, J. Bai, F. Yang, and X. Bai, "Receptive-field and direction induced attention network for infrared dim small target detection with a large-scale dataset IRDST," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5000513.

[16] Y. Dai, X. Li, F. Zhou, Y. Qian, Y. Chen, and J. Yang, "One-stage cascade refinement networks for infrared small target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5000917.

[17] J. Du et al., "A spatial–temporal feature-based detection framework for infrared dim small target," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 3000412.

[18] M. Wan, X. Ye, X. Zhang, Y. Xu, G. Gu, and Q. Chen, "Infrared small target tracking via Gaussian curvature-based compressive convolution feature extraction," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 7000905.

[19] Z. Wang, T. Zang, Z. Fu, H. Yang, and W. Du, "RLPGB-Net: Reinforcement learning of feature fusion and global context boundary attention for infrared dim small target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5003615.

[20] M. M. Hadhoud and D. W. Thomas, "The two-dimensional adaptive LMS (TDLMS) algorithm," *IEEE Trans. Circuits Syst.*, vol. CS-35, no. 5, pp. 485–494, May 1988.

[21] D. Konstantin and D. Zosso, "Two-dimensional variational mode decomposition," in *Proc. Int. Workshop Energy Minimization Methods Comput. Vis. Pattern Recognit.*, Hong Kong, 2015, pp. 13–16.

[22] J. Han, S. Moradi, I. Faramarzi, C. Liu, H. Zhang, and Q. Zhao, "A local contrast method for infrared small-target detection utilizing a tri-layer window," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 10, pp. 1822–1826, Oct. 2020.

[23] Y. Dai and Y. Wu, "Reweighted infrared patch-tensor model with both nonlocal and local priors for single-frame small target detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 8, pp. 3752–3767, Aug. 2017.

[24] L. Zhang and Z. Peng, "Infrared small target detection based on partial sum of the tensor nuclear norm," *Remote Sens.*, vol. 11, no. 4, p. 382, Feb. 2019.

[25] Y. Dai, Y. Wu, F. Zhou, and K. Barnard, "Attentional local contrast networks for infrared small target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 11, pp. 9813–9824, Nov. 2021.

[26] M. Zhang, R. Zhang, Y. Yang, H. Bai, J. Zhang, and J. Guo, "ISNet: Shape matters for infrared small target detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 877–886.

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

[28] Q. Hou, Z. Wang, F. Tan, Y. Zhao, H. Zheng, and W. Zhang, "RISTDnet: Robust infrared small target detection network," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 7000805.

[29] G. Chen, W. Wang, and S. Tan, "IRSTFormer: A hierarchical vision transformer for infrared small target detection," *Remote Sens.*, vol. 14, no. 14, p. 3258, Jul. 2022.

[30] R. Li and Y. Shen, "YOLOSR-IST: A deep learning method for small target detection in infrared remote sensing images based on super-resolution and YOLO," *Signal Process.*, vol. 208, Jul. 2023, Art. no. 108962.

[31] X. Zhou, L. Jiang, C. Hu, S. Lei, T. Zhang, and X. Mou, "YOLO-SASE: An improved YOLO algorithm for the small targets detection in complex backgrounds," *Sensors*, vol. 22, no. 12, p. 4600, Jun. 2022.

[32] S. Yao, Q. Zhu, T. Zhang, W. Cui, and P. Yan, "Infrared image small-target detection based on improved FCOS and spatio-temporal features," *Electronics*, vol. 11, no. 6, p. 933, Mar. 2022.

[33] M. Ju, J. Luo, G. Liu, and H. Luo, "ISTDet: An efficient end-to-end neural network for infrared small target detection," *Infr. Phys. Technol.*, vol. 114, May 2021, Art. no. 103659.

[34] C.-Y. Wang, A. Bochkovskiy, and H.-Y. Mark Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," 2022, *arXiv:2207.02696*.

[35] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2980–2988.

[36] J. Wang, C. Xu, W. Yang, and L. Yu, "A normalized Gaussian Wasserstein distance for tiny object detection," 2021, *arXiv:2110.13389*.

[37] X. Dai et al., "Dynamic head: Unifying object detection heads with attentions," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2021, pp. 7373–7382.

[38] J. Dai et al., "Deformable convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 764–773.

[39] T. Zhang, S. Cao, T. Pu, and Z. Peng, "AGPCNet: Attention-guided pyramid context networks for infrared small target detection," 2021, *arXiv:2111.03580*.

Bo Yang was born in 1995. He received the M.S. degree from Chongqing Jiaotong University, Chongqing, China, in 2018. He is currently pursuing the Ph.D. degree with the College of Mechanical and Vehicle Engineering, State Key Laboratory of Mechanical Transmission, Chongqing University, Chongqing.

His research interests include deep learning, target detection, intelligent unmanned system, and target tracking.

Xinyu Zhang received the B.S. degree from the School of Mechanical Engineering, Tiangong University of China, Tianjin, China, in 2021. He is currently pursuing the M.S. degree with the School of Mechanical and Vehicle Engineering, Chongqing University, Chongqing, China.

His research interests include maneuver trajectory prediction, intelligent unmanned system, and intention recognition.

Jian Zhang received the B.S. degree from the School of Civil Engineering, Chongqing Jiaotong University, Chongqing, China, in 2014, and the M.S. degree from the School of Aerospace Engineering and Applied Mechanics, Tongji University, Shanghai, China, in 2017. He is currently pursuing the Ph.D. degree with the College of Mechanical and Vehicle Engineering, Chongqing University, Chongqing.

From 2017 to 2018, he was an Assistant Engineer with the State Key Laboratory of Vehicle NVH and Safety Technology, Chongqing. His research interests include nonlinear dynamics, nonlinear control, distributed parameter system, and flexible robotics.

Jun Luo received the B.S. and M.S. degrees in mechanical engineering from Henan Polytechnic University, Jiaozuo, China, in 1994 and 1997, respectively, and the Ph.D. degree in mechanical engineering from the Research Institute of Robotics, Shanghai Jiao Tong University, Shanghai, China, in 2000.

He is currently a Professor with the State Key Laboratory of Mechanical Transmissions, Chongqing University, Chongqing, China. His current research interests include artificial intelligence, sensing technology, intelligent unmanned system, and special robotics.

Mingliang Zhou (Member, IEEE) received the Ph.D. degree in computer science from Beihang University, Beijing, China, in 2017.

He was a Post-Doctoral Researcher with the Department of Computer Science, City University of Hong Kong, Hong Kong, from September 2017 to September 2019. He is currently a Lecturer with the School of Computer Science, Chongqing University, Chongqing, China. He is also a Post-Doctoral Researcher with the State Key Laboratory of Internet of Things for Smart City, University of Macau, Macau, China. His research interests include image and video coding, perceptual image processing, multimedia signal processing, rate control, multimedia communication, machine learning, and optimization.

Yangjun Pi received the B.Eng. degree in mechatronic engineering and the Ph.D. degree in mechanical engineering from Zhejiang University, Hangzhou, China, in 2005 and 2010, respectively.

He is currently a Professor with the State Key Laboratory of Mechanical Transmissions and the College of Mechanical and Vehicle Engineering, Chongqing University, Chongqing, China. His research interests include control of distributed parameter systems, intelligent unmanned system, and vibration control.