Meta-learning based infrared ship object detection model for generalization to unknown domains

Hui Feng, Wuhan University of Technology

Wei Tang, Wuhan University of Technology

Haixiang Xu, Wuhan University of Technology

Chengxin Jiang, Wuhan University of Technology

Shuzhi Sam Ge, National University of Singapore

Jianhua He, University of Essex

Meta-Learning based Infrared Ship Object Detection Model for Generalization to Unknown Domains

Hui Feng[a,b], Wei Tang[a,b], Haixiang Xu[a,b,*], Chengxin Jiang[a,b], Shuzhi Sam Ge[c], Jianhua He[d]

[a]*Key Laboratory of High Performance Ship Technology (Wuhan University of Technology), Ministry of Education, Wuhan, 430063, China*
[b]*School of Naval Architecture, Ocean and Energy Power Engineering, Wuhan University of Technology, Wuhan, 430063, China*
[c]*Department of Electrical and Computer Engineering, National University of Singapore, Singapore 117576, Singapore*
[d]*School of Computer Science and Electronic Engineering, University of Essex, CO4 3SQ, United Kingdom*

Abstract

Infrared images exhibit considerable variations in probability distributions, stemming from the utilization of distinct infrared sensors and the influence of diverse environmental conditions. The variations pose great challenges for deep learning models to detect ship objects and adapt to unseen maritime environments. To address the domain shift problem, we propose an end-to-end infrared ship object detection model based on meta-learning neural network to improve domain adaptation for target domain where data is not available at training phase. Different from existing domain generalization methods, the novelty of our model lies in the effective exploition of meta-learning and domain adaptation, ensuring that the extracted domain-independent features are meaningful and domain-invariant at the semantic level. Firstly, a double gradient-based meta-learning algorithm is designed to solve the common optimal descent direction between different domains through two gradient updates in the inner and outer loops. The algorithm enables extraction of domain-invariant features from the pseudo-source and pseudo-target domain data. Secondly, a domain discriminator with dynamic-weighted gradient reversal layer (DWGRL) is designed to accurately classify domain-invariant features and provide additional global supervision information. Finally, a multi-scale feature aggregation method is proposed to improve the extraction of multi-scale domain-invariant features. It can effectively fuse local features at different scales and global features of targets. Extensive experimental results conducted in real nighttime water surface scenes demonstrate that the proposed model achieves very high detection accuracy on target domain data, even no target domain data was used during the training phase. Compared to the existing methods, our method not only improves the detection accuracy of infrared ships by 18%, but also exhibits the smallest standard deviation with a value of 0.93, indicating its superior generalization performance.

Keywords: Infrared object detection, Meta-learning, Domain discriminator, Intelligent ship

1. Introduction

In the context of intelligent ship visual perception, object detection holds immense importance as it plays a critical role in ensuring safe navigation[1–5]. Enabling nighttime navigation for intelligent ships has become a significant research focus, resulting to the growing attention on infrared image-based water surface object detection. The choice of infrared sensor types and factors like varying atmospheric temperature and illumination conditions in maritime environments, significantly impact image quality and subsequently affect object detection performance. Therefore, investigating and understanding the effects of different infrared sensors and the unique characteristics of maritime environments are essential for enhancing the reliability and efficacy of infrared-based object detection methods in these scenarios.

*Corresponding author
Email addresses: feng@whut.edu.cn (Hui Feng), Kingertw@163.com (Wei Tang), qukaiyang_whut@163.com (Haixiang Xu), xiiiaoxin@whut.edu.cn (Chengxin Jiang), samge@nus.edu.sg (Shuzhi Sam Ge), j.he@essex.ac.uk (Jianhua He)

Recently, infrared object detection algorithms based on convolutional neural networks (CNNs) have made remarkable progress[6, 7]. These methods have shown great promise in detecting infrared ship targets, benefiting from their robust representational capabilities and ability to extract intricate features. However, CNN-based methods used in infrared image object detection often encounter the challenge of domain shift, due to the substantial data requirements. The object detection algorithm trained on the source domain often suffers significant performance degradation when applied to the target domain due to its limited generalization ability to overcome the impact of domain shift[8]. Domain shift is primarily attributed to two factors. Firstly, non-uniformity differences between infrared sensors, resulting from production process limitations, lead to distribution variations among infrared image data[9]. Secondly, changing lighting conditions and sea temperature cause significant differences in the characteristics of ship objects. Although collecting images from each infrared sensor or at different times can alleviate domain shift problems, it often requires significant time and annotation costs. Moreover, in certain scenarios, acquiring data from the target domain may prove challenging. It is challenging to solve the domain shift issue and enhance model generalization capability.

Domain adaptation (DA) and domain generalization (DG) are two general approaches to address domain shif issues. DA aims to align the source domain and target domain in the feature space, thereby extracting robust features[10–13]. Although DA offers a solution to the time-consuming task of labeling data, it still requires a significant amount of labeled data for successful training[14, 15], which may not always be feasible in real-world scenarios. As a result, there has been a growing interest in the field of domain generalization (DG) in recent years[15–18]. However, in DG, the absence of target domain data limits the alignment to be conducted only between the source domains. This raises concerns about potential overfitting of the model on the source domains[15, 19, 20]. Robust representation learning is achieved only when the source domain exhibits shifts, which are simulated using both the source and meta-target domains. But this problem will be exacerbated if the source domain lacks diversity. Moreover, traditional methods used to address domain shift issues in visible images may not work well for infrared ship object detection algorithms.

This paper proposes an end-to-end infrared ship object detection algorithm based on meta-learning and domain adaptation (IS-MLDA) to address the domain shift problem between the source and target domains. To the best of our knowledge, this study represents the first attempt to combine meta-learning and domain discriminator techniques in the field of infrared ship object detection. Firstly, a meta-learning method based on double gradient descent is designed to minimize the meta-training domain loss and optimize the meta-test domain loss. Through the meta-learning process, the model effectively extracts domain-invariant features, enabling improved performance in both the source and target domains. Secondly, a domain discriminator with a DWGRL is designed to provide the model with more global supervision information, preventing overfitting on the source domain data. Lastly, a multi-scale feature fusion module is designed to aggregate local features at different receptive fields and global features, refining the fused features and reducing the aliasing effect introduced by the fusion process, ultimately enhancing the model's feature representation capability.

The contributions of this paper are summarized as follows:

(1) The proposed IS-MLDA enables end-to-end detection of infrared ship targets, eliminating the need for tedious image preprocessing in infrared ship object detection. By training the model solely on source domain data, it achieves accurate target detection in previously unseen domains. The designed training process effectively extracts domain-invariant features, addressing the problem of poor generalization ability and detection performance of object detection networks in the target domain.

(2) For the first time, we apply gradient-based meta-learning algorithms and domain adaptation methods to the field of infrared ship object detection. We validate the effectiveness of combining domain generalization and domain adaptation for solving the domain shift problem.

(3) We introduce a multi-scale feature fusion module to aggregate local and global features, which enhanced the model's feature representation ability.

(4) We develop new datasets for ship detection and domain adaptation, simulating the domain shift phenomenon using infrared ship images from different infrared sensors in real-world scenarios. Experimental results demonstrate that our method outperforms existing object detection algorithms in terms of object detection and generalization performance, providing robust technical support for the nighttime navigation safety of intelligent ships.

The remaining sections of this paper are organized as follows. Section 2 provides an overview of the research status in ship infrared image object detection and domain shift. Section 3 presents the training strategy and detailed network structure of the proposed IS-MLDA model. In Section 4, we conduct experimental comparisons and analyses

between IS-MLDA and other object detection algorithms. Finally, Section 5 summarizes the proposed approach and concludes the paper.

2. Realated Work

2.1. Infrared object detection

With the advancements in infrared imaging technology, infrared sensors have demonstrated strong penetration capability and excellent concealment in challenging environments such as nighttime, rain, fog, and smoke. Therefore, researchers have paid significant attention to object detection based on infrared images. Traditional infrared object detection methods, such as spatial filtering[21] and frequency domain filtering[22], while simple to implement, often perform poorly in complex sea conditions and are susceptible to environmental interference.

Significant progress has been recently made in infrared object detection algorithms based on convolutional neural networks (CNNs)[6, 7, 23]. These methods can be categorized into two groups based on their approach to generating candidate regions and classifying objects: two-stage and single-stage methods[2, 24, 25]. Single -stage algorithms are well-known for their simplicity, making them suitable for implementation and deployment on embedded devices and mobile platforms. Zhao et al.[7] utilized generative adversarial networks to generate detection images that retain only the target, converting infrared target detection into an image-to-image conversion task, and overcoming the influence of image noise on detection accuracy. Miao et al.[26] designed Fourier transform and lightweight CNN for feature cascading decision-making, gradually eliminating false positives and ensuring high detection recall rates. Song et al.[27] introduced the HaarConv module to enhance the feature extraction capability of the model backbone network. Additionally, Du et al.[28] applied transfer learning techniques, processing visible light images on infrared datasets, which effectively suppressed complex backgrounds and reduced false detection rates.

CNN-based methods have exhibited exceptional performance and achieved high accuracy in ship detection. However, the effectiveness of these detection methods is primarily limited to datasets with similar distributions. Therefore, deep learning methods may not perform effectively when confronted with training and test sets that exhibit different data distributions[15, 29]. The main reason for this situation is the presence of domain shift between the training set and the test set[18]. In real-world scenarios, the detection accuracy of a model trained on the source domain is significantly reduced when directly applied to the target domain[8, 30]. This phenomenon can be attributed to variations in infrared sensors, such as differences in pixel size and imaging angle, as well as the influence of ambient temperature on the imaging process, leading to diverse probability distributions among the infrared images. Existing CNN-based methods are highly susceptible to domain shift, resulting in overfitting to the source domain[31–33].

While collecting extensive data from each new domain can help mitigate the challenges associated with domain shift, this approach necessitates substantial investments in terms of time and resources for annotation. Additionally, acquiring data from the target domain may prove challenging in some instances. Hence, it is crucial to address the impact of distribution differences between the training and test sets arising from variations in infrared sensor imaging processes, photosensitive elements, viewing angles, and maritime environments, which can significantly affect the efficacy of target detection algorithms.

2.2. Domain Adatation and Domain Generalization

Domain shift is a common problem in deep learning, referring to the performance degradation of a model trained on a source domain when applied to a target domain with different statistical data distribution [16, 17, 34]. Domain shift can have significant impact on deep learning models as access to data that accurately represents the testing scenario may not be available. Recently the approaches of DA and DG to address domain shift issue gained increasing attention.

Domain Adatation: The main concept of DA is domain alignment across domains, which requires access to the source domain as well as a small amount of target domain data during the training process[14]. Most domain adaptation methods based on domain alignment are designed to either minimize the distribution difference between the source and target domain data or maximize the domain classifier loss. Xia et al.[35] designed an adaptive adversarial network A2-Net based on the passive domain adaptation method, which can effectively classify images between different domains without source data. By using the second-order statistic distance of the features extracted by the model, Sun et al.[36] designed a new loss function and regularization term to minimize the feature distance between the

source and target domains. Zhang et al.[37] proposed an unsupervised domain adaptation method based on boundary difference, which introduces a gradient inversion layer for domain adaptation image classification.To avoid the degradation of data features during distribution alignment, discriminative invariant alignment (DIA) [38] enriches the knowledge matrix by combining the class discriminative information of the source domain and local data structure information of the target domain into a new framework. By introducing the maximum margin criterion of the source domain, the classification boundaries are expanded. The negative transfer of data feature outliers can occur during domain alignment, and to address this issue, Lu et al. proposed Weighted Correlation Embedding Learning (WCEL) [39] and guided discrimination and correlation subspace learning (GDCSL) [40]. GDCSL fully considers domain-invariant and discriminative features, avoiding the problem of class center shift caused by domain alignment process. In addition, GDCSL uses relevant features from source and target domains to enhance the classification ability of the model. DA typically involves aligning the source and target domains by fine-tuning the model with a limited amount of target domain data. However, this approach assumes that the target domain data follows the same probability distribution as the source domain data, which may not always hold true in practical scenarios. Our objective is to develop a detection method that can achieve satisfactory results in the new target domain without the need for fine-tuning with target domain data. Moreover, while domain adaptation has been extensively studied in the context of visible image classification, there is a lack of research on infrared ship target detection. The absence of texture information and the complex and variable nature of texture features make the distribution of ship targets in infrared images significantly distinct. As a result, enhancing the detection accuracy of infrared ship targets in an unknown domain poses a more formidable challenge.

Domain Generalization: DG aims to learn a model that can be trained using data from one or more source domains that related but distinct and subsequently applied to any out of distribution (OOD) unseen target domains without requiring any form of adaptation[18, 20]. Most existing DG methods focus on training a model that is invariant to the source domain and aligns the distribution of the source domain in the feature space[34, 41, 42]. The DG approach, which is rooted in meta-learning, mitigates the issue of overfitting that arises from the source domain. This method leverages the meta-test set to simulate the presence of an unseen domain[31, 43]. MLDG (meta-learning for domain generalization)[34] is the first study of applying meta-learning algorithms to DG, which randomly divides multiple source domains into pseudo-source and pseudo-target domains, and then uses meta-learning algorithms to optimize the loss functions of pseudo-source and pseudo-target domains to guide the model to learn features with strong generalization ability. In MLDG, the key role of meta-learning is to use the training dataset to simulate domain shift in real-world scenarios, thus generating a model that can be directly applied to the testing domain (target domain).[44] proposed Metareg to learn a regularizer that can achieve good cross-domain generalization in the meta-learning algorithms to address the problem of domain shift. The key idea of DG methods is to reduce the difference between the source and target domains in the feature space during the training process.

The existing meta-learning-based DG method offers significant benefits in zero-shot learning. However, it is important to note that in the absence of diversity in the source domain data, it is difficult to determine that the extracted features are domain-invariant, and there is still a risk of overfitting. Therefore, the method designed in this paper provides more global supervision information for the meta-learning process and avoids overfitting the model on the source domain.

3. Proposed Method

We propose an infrared ship object detection algorithm based on meta-learning and domain adaptation (IS-MLDA) to address the lack of generalization ability of existing infrared ship detection algorithms in unknown domains. Firstly, meta-learning is employed to simulate domain shift between different domains during model training. By adjusting the coordinated descent of the gradients in the pseudo-source and pseudo-target domains, the model is guided to learn domain-invariant features while filtering out domain-related information. Secondly, a novel domain discriminator with a DWGRL is utilized to classify the domain-invariant features in the feature space and reduce model overfitting to the source domain data. Finally, a multi-scale feature aggregation structure is proposed to increase the model's receptive field and improve detection performance for multi-scale ship targets.

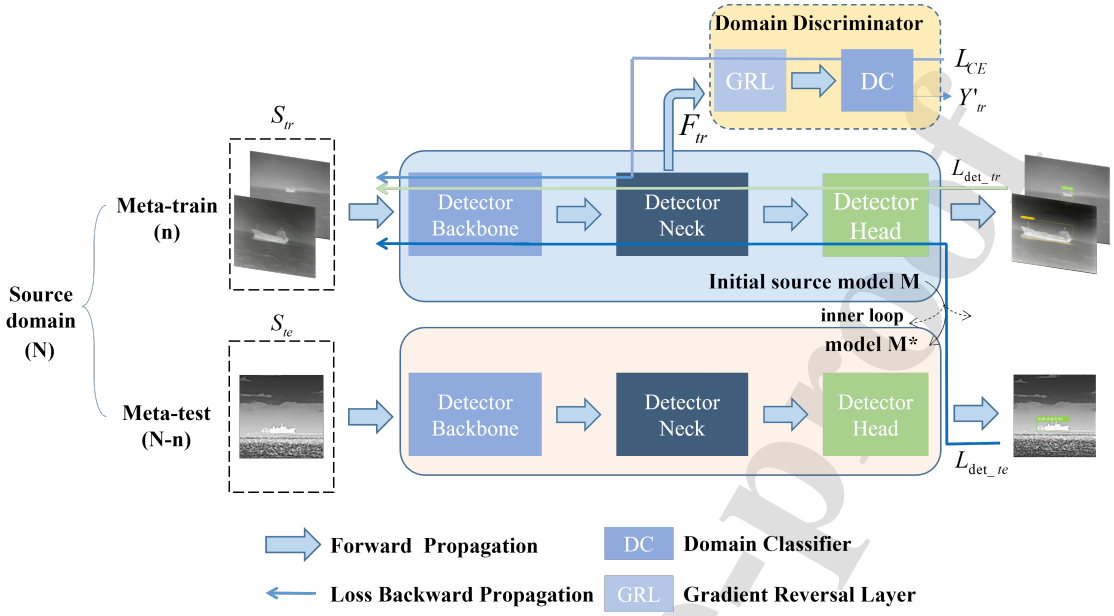


Fig. 1. The pipeline of the IS-MLDA.The forward arrow represents the data forward propagation process, and the backward arrow indicates the direction of gradient propagation. IS-MLDA divides each iteration process into three stages: meta-train stage, meta-test stage, and total loss backpropagation stage. Before training, N source domain datasets need to be partitioned into n meta-train domains and N-n meta-test domains. The inner loop refers to the process of updating during each iteration. S_{tr} represents the meta-train set, S_{te} represents the meta-test set, F_{tr} denotes the output features of the detector's neck layer, L_{CE} represents the domain discriminator loss, L_{det_tr} represents the target detection loss for the meta-train set, and L_{det_te} represents the target detection loss for the meta-test set.

3.1. The training process of IS-MLDA

The model training process is depicted in Fig. 1. The primary objective is to train the model on the source domain to learn domain-invariant features across different domains, enabling the model to generalize well when tested on the target domain. Inspired by MLDG[34], we have designed a training process based on double gradient descent. Next the training process is presented in detail.

We assume that N source domains are represented by $D = \{d_1, d_2, ..., d_n\}$, sharing the same detection task but having inconsistent data distribution of the images. The initial source model is represented as M. In each iteration, n domains are randomly selected from N source domains as the **meta-train** set (pseudo-source domain), which is denoted by S_{tr}, and the remaining N-n domains form the **meta-test** set (pseudo-target domain), which is denoted by S_{te}.

Meta-train: Initially, infrared images of S_{tr} are fed into the backbone network and the neck layer to extract features, denoted by F_{tr}. Subsequently, the features S_{tr} are fed into the domain discriminator and the head of detector to obtain domain classification label Y'_{tr} and predicted object detection result, respectively. The domain discriminator loss L_{CE} and the object detection loss L_{det_tr} are calculated separately. Finally, the total loss function over the meta-train set is computed by weighting the sum of L_{CE} and L_{det_tr} . The initial source model M is then updated as M* through the first gradient descent (i.e., inner loop update).

Meta-test: During the meta testing process, infrared images of S_{te} are fed into M* to obtain detection results, then we can calculate the object detection loss L_{det_te}. The purpose of meta-test is to simulate domain shift in real-world scenarios and evaluate the detection performance of the model in unknown domains.

Backpropagation of the total loss: After obtaining L_{CE}, L_{det_tr}, L_{det_te}, the total loss is computed by summing them up, and the gradient of the total loss is calculated to update the model parameters, which is the second gradient descent. It is noted that the gradient of the total loss is computed based on the initial source model parameters.

For ease of understanding the training process of IS-MLDA, we have correspondingly simplified the propagation

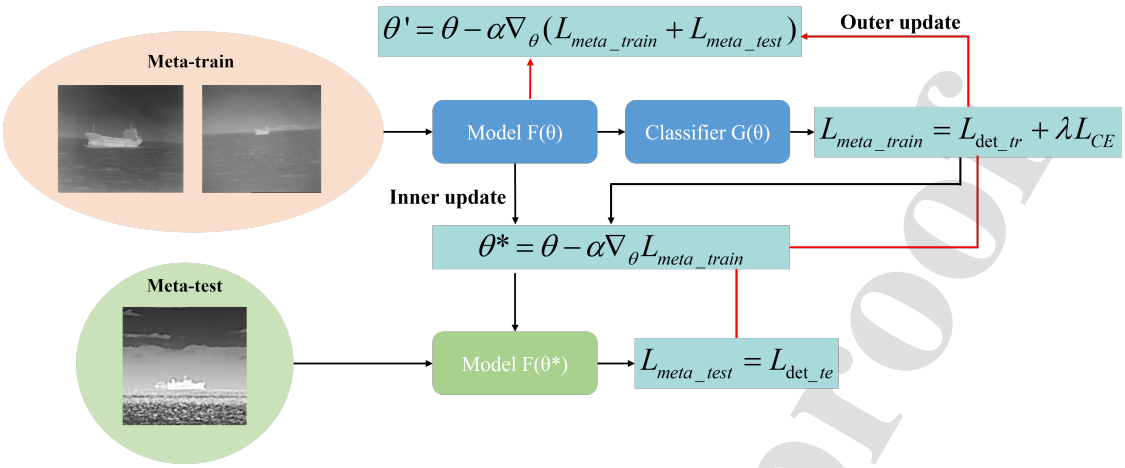


Fig. 2. The parameter update process of IS-MLDA. The black line in the figure represents the inner loop update process, and the red line represents the outer loop update process. θ represents the initial parameters of the source model, θ^* represents the model parameters after the inner update, ∇_θ represents the gradient based on the parameters θ, θ' represents the model parameters after the outer update, L_{meta_train} represents the total loss during the meta-training process and L_{meta_test} represents the loss during the meta-testing process.

of the loss functions in Fig. 1. As shown in Fig. 2, after the meta-train process, the weighted loss of the target detection loss and domain classifier loss is calculated and denoted as L_{meta_train}, and the model parameters are updated based on the weighted loss. In the meta-test phase, based on the updated model, the target detection loss of the meta-test set is calculated and denoted as L_{meta_test}. In the backpropagation stage of the loss function, the total loss function of the meta-train phase and the meta-test phase is calculated, and then a gradient based on the initial model parameters is calculated using the total loss function for parameter updating.

Due to the variations in data distribution among infrared ship images, the detection accuracy of the model is significantly impacted, leading to reduced performance. Employing meta-learning training strategies to guide the model in learning robust and generalizable ship features is essential for improving its detection accuracy across diverse image domains.

3.2. Network Architecture

The model architecture of IS-MLDA is illustrated in Fig. 3. During training of model, a domain discriminator after the neck layer of the detector is introduced to classify the domain-invariant features extracted by the model. This domain discriminator provides additional global supervision information to the backbone network and neck layer during training, effectively mitigating the risk of overfitting to the source domain. Importantly, the domain discriminator is utilized solely during the training phase to optimize the domain-agnostic features extracted by the model and does not introduce additional time cost during the testing phase. To enhance the detection accuracy of the model for objects at varying scales, we develop a multi-scale feature aggregation structure (MFA-s). The MFA-s is designed to aggregate local features at different scales and incorporate global features, refining the fused features in the channel dimension to improve the model's feature representation capability.

3.2.1. Dynamic-weighted gradient reversal layer

Due to variations in the distribution of ship targets among different infrared images, determining the domain-invariance of features extracted by meta-learning poses a challenge. Moreover, relying solely on meta-learning algorithms may not provide comprehensive global supervision for the model, thus increasing the risk of overfitting to the source domain data. Inspired by the domain adaptation methods[35], we propose a domain discriminator integrated with the meta-learning algorithm after the neck layer of the detector. This integration aims to optimize domain-invariant features and enhance the model's generalization across different domains. The architecture of the proposed

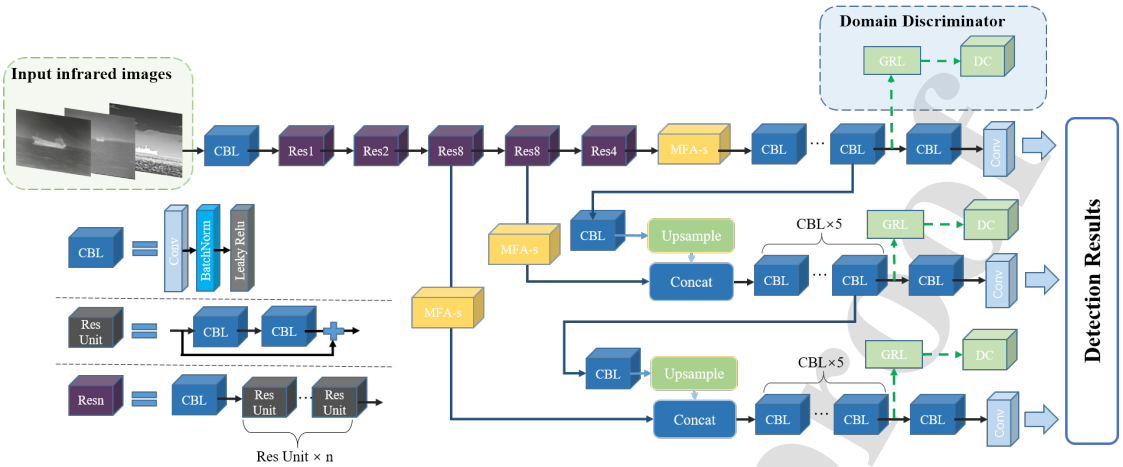
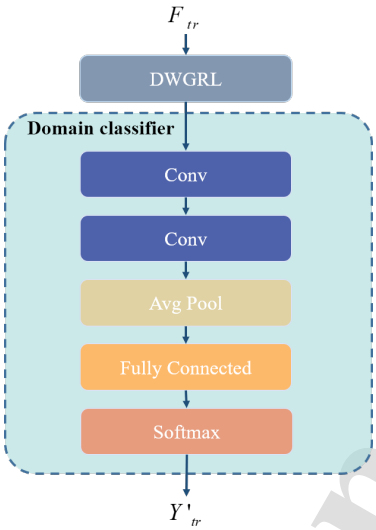
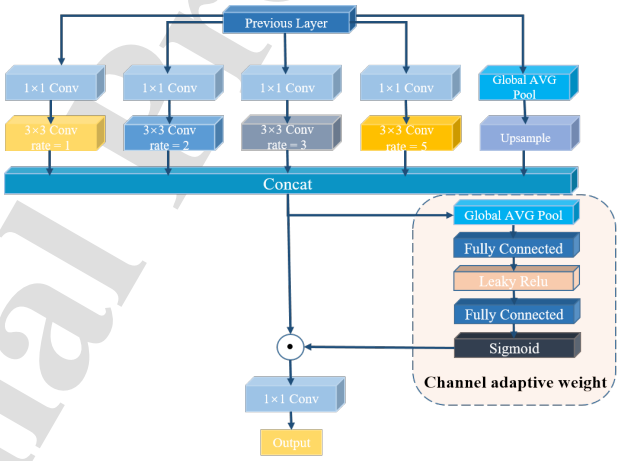


Fig. 3. The network structure of IS-MLDA. IS-MLDA's baseline model adopts the YOLOv3 object detection model, where CBL represents a convolutional block, Res Unit represents a residual unit, Res *n* represents a combination of a convolutional block and *n* residual units, and MFA-s represents the multi-scale feature aggregation structure. It is noteworthy that the domain discriminator is introduced only during the training phase and does not impact the efficiency of the inference process.



(a) The network structure of Domain Discriminator.



(b) The network structure of MFA-s.

Fig. 4. The network structure of Domain Discriminator and Multi-scale Feature Aggregation (MFA-s). (a) illustrates the structural details of the domain discriminator based on DWGRL, where DWGRL is designed to control the gradient propagation intensity of the domain classifier. (b) presents the detailed structure of MFA-s. We utilize five parallel dilated convolution branches to correlate local and global features in the feature map. We compute response weights for infrared ship targets in different channels, assigning larger weights to channels with greater target responses to enhance the features of infrared targets.

domain discriminator, as shown in Fig. 4(a), consists of a dynamic-weighted gradient reversal layer (DWGRL) followed by a domain classifier with 3×3 convolutional layers and a fully connected layer. The GRL[45] helps models learn more robust and domain-invariant features by making it harder for the domain classifier to accurately identify the source of input data, improving model generalization ability. During the forward propagation, the GRL acts as an

identity mapping, preserving the input values without any modifications. Mathematically, this can be represented as:

$$R_\lambda(x) = x \tag{1}$$

where x is the output feature of the neck layer. During the backpropagation process, the GRL reverses the sign of the gradient of the domain classification loss, effectively maximizing the domain discriminator loss and encouraging the model to learn domain-invariant features. This can be mathematically expressed as:

$$\frac{dR_\lambda(x)}{dx} = -\lambda I \tag{2}$$

where I is the gradient of the domain classifier loss, the weight coefficient λ is used to control the strength of gradient reversal.

Traditionally, the GRL sets the weight coefficient λ to a fixed value, such as 1. However, at the initial training stage of the model, the domain classifier may struggle to accurately classify domain-invariant features, potentially leading to incorrect guidance for the feature extraction process. To address this issue, we propose a dynamic-weighted gradient reversal layer (DWGRL).

We propose a dynamic weighting mechanism for DWGRL to process the domain classifier loss gradient during training iterations. Initially, the weight of the domain classifier loss is set to a small value, allowing the model to primarily focus on learning the main task of object detection. As the training progresses, we gradually increase the weight of the domain classifier loss using a function that depends on the current iteration number and the total iteration number. The adaptive weighting strategy can be formulated as follows:

$$\frac{dR_\lambda(x)}{dx} = -\lambda(t) I \tag{3}$$

$$\lambda(t) = 5 \cdot \frac{e^t - e^{-t}}{e^t + e^{-t}} \tag{4}$$

where $t \in [0, 1]$, it represents the ratio of the current iteration number to the total number of iterations. $\lambda(t)$ is a function that controls the dynamic adjustment of the weight. By dynamically adjusting the weight of the domain classifier loss during the training process, the DWGRL method encourages the model to prioritize domain-invariant feature extraction and domain adaptation as it gains experience from the training data. This approach enhances the stability of the model training process and facilitates better generalization to different domains, leading to improved detection performance in infrared ship object detection tasks.

3.2.2. Multi-scale Feature Aggregation structure

Various single-stage object detection algorithms, including YOLOX[46], have recently been proposed to achieve high-precision detection of targets in visible images. However, their computational requirements make them less suitable for deployment on edge devices with limited computing power, such as those used in intelligent ships. To address the need for real-time and efficient detection on edge devices, we evaluated different single-stage object detection algorithms. Among them, YOLOv3[47] emerged as a desired model that strikes an excellent balance between speed and accuracy, which is selected YOLOv3 as the baseline model for the infrared ship target detection. In the neck of the YOLOv3 model, the FPN structure[48] is used to enhance the fusion of multi-scale features, aiming to improve the detection performance of the model for targets of different scales. However, as illustrated in Fig. 5, the detection performance may severely degrade due to the significant scale variation and the scarcity of detailed features of ship targets in infrared images. To address this issue and bolster the model's capability to effectively capture multi-scale features, we devise the MFA-s module, which is integrated after each of the three effective feature layers. The primary objective of the MFA-s module is to aggregate contextual information from multi-scale features and expand the receptive field. It can enhance the model's ability to detect targets of various scales, as illustrated in Fig. 4(b). The MFA-s module is designed with several specific objectives. Firstly, we leverage dilated convolutions with different dilation rates and a 3×3 kernel size to effectively associate local features from the preceding layer with a wider field of view, significantly improving the model's feature extraction ability for ship targets of various scales. To reduce the parameter size, we apply a 1×1 convolutional layer for dimensionality reduction before employing the dilated

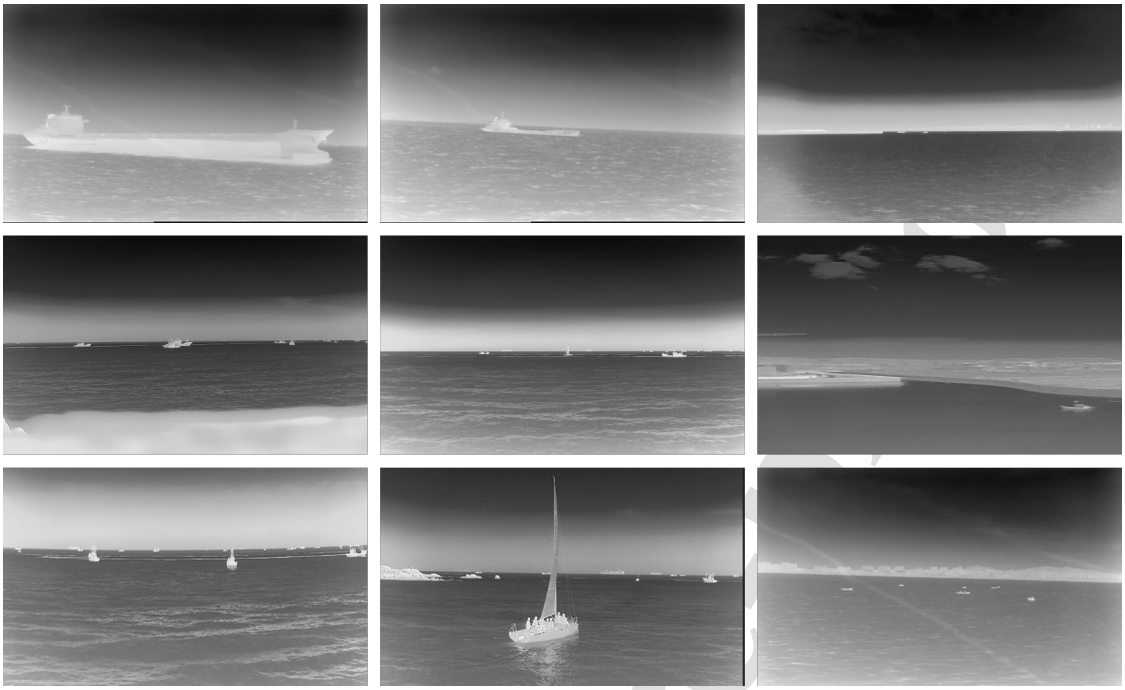


Fig. 5. Multi-scale objects in the dataset.

convolutions. In addition, considering the distinct responses of infrared ship target features in different channels and their correlation, features from different branches along the channel dimension are concatenated. Subsequently, we compute the response weight of the target in each channel using global average pooling and fully connected layers. This step helps to enhance infrared target features and mitigates the aliasing effect introduced during the fusion process. Finally, the fused feature map is multiplied by the computed response weight channel-wise, enabling the model to prioritize the extraction of effective features.

3.3. Loss Function

As outlined in Section 3.1, the IS-MLDA training process consists of two distinct stages: meta-training and meta-testing . During the meta-training stage, we tackle both the domain classification task and the object detection task. Let L_{CE} represent the loss of the domain discriminator and $L_{\text{det\_}tr}$ denote the loss of the object detection task. For the domain classification task, we optimize the domain discriminator using the cross-entropy loss function:

$$L_{CE}(o, class) = \sum_{b=1}^{N} r_b \left(-o_{class} + \log \sum_{j=1}^{N} \exp(o_j) \right) \tag{5}$$

where, o represents the result of the domain discriminator, represents the domain class label, and $class$ is the indicator function. The loss function for object detection during the meta-train stage is represented as follows:

$$L_{\text{det\_}tr} = \lambda_{box} L_{box} + \lambda_{conf} L_{conf} + \lambda_{cls} L_{cls} \tag{6}$$

where λ_{box} , λ_{conf} and λ_{cls} are weight coefficients of box coordinate loss, confidence loss and target classification loss, respectively.

The total loss function of the meta-train stage can be represented as:

$$L_{meta\_train} = \lambda_{CE} L_{CE} + L_{\text{det\_}tr} \tag{7}$$

where λ_{CE} is a weight coefficient of domain discriminator.

During the meta-testing stage, it tests the infrared images in the meta-test set to evaluate the generalization performance of model in unseen domains. The total loss function during the meta-training stage is represented by the loss function for object detection, which is given by:

$$L_{meta\_test} = L_{\det\_te} = \lambda_{box}L_{box} + \lambda_{conf}L_{conf} + \lambda_{cls}L_{cls} \tag{8}$$

Finally, IS-MLDA calculates the loss function and performs gradient updates based on the parameters of the initial source model. The final loss function can be represented as:

$$L = L_{meta\_train} + L_{meta\_test} \tag{9}$$

4. Experimental results and analysis

4.1. Datasets and Implementation Details

The proposed IS-MLDA is trained on three distinct infrared datasets. The infrared ship dataset uses three different resolutions (384×288, 640×512, and 1280×1024) of infrared sensors to capture ship targets at sea and ports in different scenes, perspectives, and time periods. Each ship in the dataset has been annotated with labels according to its target category. The label categories include liner, bulk carrier, warship, sailboat, canoe, container ship, and fishing boat. Although these images feature the same ship class, their data distributions vary significantly. To simulate domain shift issues in real-world scenarios, we partitioned the dataset into three domains based on differences in data distribution and assigned domain labels (domain_id) to them. Specifically, domain_id=[0, 1, 2]. To construct cross-domain object detection tasks to verify the generalization ability of model, we choose two datasets as source domains and one as a target domain (AB→C, AC→B, BC→A). We then divide the selected source domain datasets into training sets and test sets at a ratio of 8:2. The specific information of the dataset is shown in Table 1.

As shown in Fig. 6(a), we utilize t-Distributed Stochastic Neighbor Embedding (t-SNE) to reduce the dimension of data distribution of distinct infrared sensors for visualization, in order to better illustrate the distribution disparities between different infrared images. It is clear that the distribution of data from different infrared sensors differs significantly. Moreover, we also observed considerable changes in the distribution of data from the same infrared sensor due to the influence of environmental factors. By leveraging t-SNE for data visualization, we gain valuable insights into the variations and disparities in data distribution, which further motivate the need for our proposed IS-MLDA approach. It enables us to effectively address the challenges associated with domain shift and improve the generalization performance of the model across different infrared domains.

The training of IS-MLDA takes a total of 200 epochs and adopts Adam optimizer. The batch size is 8 and the initial learning rate is 0.0001. The weight coefficient λ_{CE} is 0.5, and λ_{box}, $\lambda_{conf}$and λ_{cls} are all equal to 1. The method in this paper is implemented based on the Pytorch[49]. All experiments in this paper are carried out on a single GPU, and the GPU model is NVIDIA GTX 2080ti. Average Precision (AP) is a general evaluation metric for object detection algorithms. Therefore, we choose the AP with an IOU threshold of 0.5 to evaluate the infrared image target detection performance of the algorithm.

4.2. Evaluation Metric

AP can be obtained from the Precision (P) and the Recall (R). The higher the average precision, the higher the object detection accuracy of the model. The calculation methods of accuracy rate $P$and recall rate R are shown in the Equation (10,11):

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

$$Recall = \frac{TP}{TP + FN} \tag{11}$$

where TP is the number of correct targets detected, FP is the number of false targets detected, and FN is the number of missed targets. Since high precision may have a large number of missed detections and high recall may have a large

Table 1: Detailed information of the infrared ship datasets with data collected by different infrared sensors.

Images source	Total number of ship classes	Source Domain	Number of image samples
Uncooled Thermal Infrared	7	A	3000
Uncooled Thermal Infrared		B	3000
Uncooled Thermal Infrared		C	3000

Table 2: Verification of the validity of DWGRL

Method	Classification Accuracy (%)	Time (ms)
Traditional GRL	94.15	1.5
DWGRL	99.68	1

number of false detections, AP should be used to evaluate the object detection performance of the algorithm. The AP calculation formula is as follows:

$$AP = \int_0^1 P(R)dR \tag{12}$$

Assuming that there are N categories of ship targets in the data, we can calculate the mean Average Precision (mAP) of all the objects to evaluate the detection performance of the model:

$$mAP = \frac{\sum_{i=1}^{N} AP_i}{N} \tag{13}$$

4.3. Quantitative and Qualitative Results

In this section, we conduct experiments to evaluate the generalization and detection performance of IS-MLDA on various domains. The experimental details are as follows:

(1) To assess IS-MLDA's generalization and detection capabilities, we partition the images from the **two source domains** into training sets and test sets in an 8:2 ratio. We then combine the test sets from both domains to form a mixed domain, while the remaining domain is treated as an unseen domain. We compare the detection accuracy and stability of IS-MLDA with state-of-the-art object detection algorithms on both the mixed domain and unseen domain data.

(2) To evaluate IS-MLDA's ability to detect multi-scale ship targets, we divide the images from the **three source domains** into training sets and test sets in an 8:2 ratio. The test sets are used as a mixed domain, and we compare the multi-scale object detection accuracy of IS-MLDA with advanced object detection methods on the mixed domain.

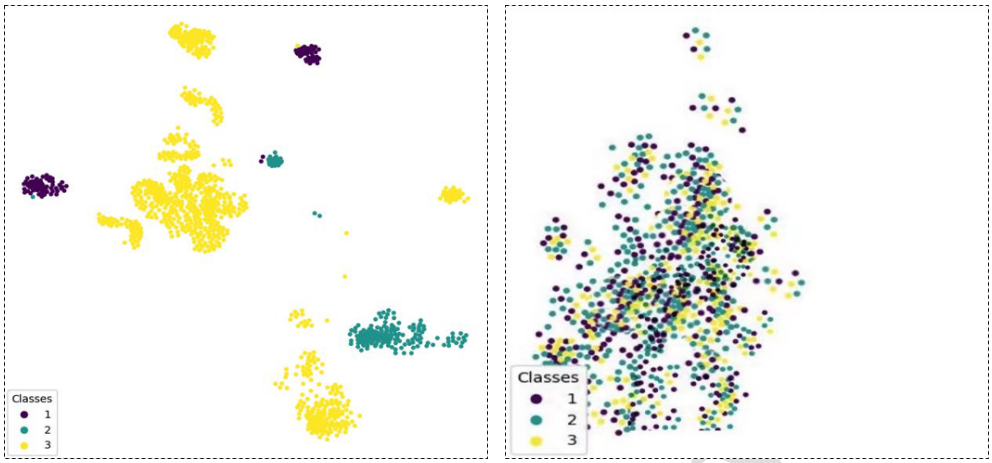
(3) Additionally, we conduct ablation experiments on IS-MLDA to investigate the contribution of each component in our method to the task of infrared ship object detection, following the experimental setup used in Experiment 1.

Furthermore, to visually demonstrate the detection performance of our method, we provide visualization of the detection results for all methods involved in the comparative experiments.

4.3.1. Verification of the validity of DWGRL

The performance of the meta-learning algorithm is heavily influenced by the accuracy of the domain classifier, as it plays a crucial role in preventing overfitting to the source domain. To rigorously assess the effectiveness of the proposed DWGRL on domain classifiers, we conducted a comprehensive classification task on images obtained from three different infrared sensors. The classification results were evaluated and compared between traditional GRL and DWGRL, as presented in Table 2.

The experimental results demonstrate that the introduction of DWGRL substantially improves the accuracy of the domain classifier by an impressive 5%, thus effectively promoting domain-invariant feature learning. Furthermore,



(a) Distribution difference of infrared images. Class 1, 2, and 3 represent domain A, B, and C respectively.

(b) The distribution of the data after feature alignment

Fig. 6. The distribution of domains. (a) Demonstrates the distribution differences of infrared images in different domains. (b) Illustrates the results after domain alignment using IS-MLDA.

DWGRL exhibits superior real-time performance. The enhanced accuracy of the domain classifier achieved through DWGRL contributes to the overall success of the IS-MLDA approach, enabling the model to generalize effectively to different domains.

4.3.2. Verification of ship object detection and generalization performance of IS-MLDA

In real-world engineering applications, variations in data distribution among infrared image domains caused by non-uniformity discrepancies across various infrared sensors and water surface temperature can pose challenges for accurate ship detection. To comprehensively evaluate the generalization performance of our proposed method on unknown domains, we conducted experiments using data collected from three different infrared sensors, which are divided into three domains denoted as A, B, and C (as detailed in Table 1).

We randomly selected two domains as the source domains, and the remaining domain was treated as the unknown domain. For evaluating detection performance, we divided the training and test sets in the two source domain images with an 8:2 ratio, and combined the test sets from both domains to form the mixed domain D. Notably, the training sets of the two source domains were not merged. We utilized mean Average Precision (mAP) with an IOU threshold of 0.5 as the evaluation metric for object detection results.

To assess the effectiveness of our proposed IS-MLDA method for infrared ship targets, we compared its performance with state-of-the-art object detection models, including Faster R-CNN[50] and SSD[51]. The detailed detection results are summarized in Table 3.

The detection results of IS-MLDA on different unseen domains are significantly better than other traditional detection methods. In the three domains of A, B, and C, the detection results of our model are higher than those of the baseline model YOLOv3(14.30%, 21.60%, 17.99%), Faster R-CNN (18.83%, 13.45%, 16.85%), and SSD (29.28%, 32.34%, 32.55%), YOLOX(2.06%, 4.09%, 1.91%). In the mixed domain D, although the detection accuracy of IS-MLDA is slightly lower than that of YOLOX, the SD reaches the minimum value of 0.99. Although our method achieved the best results, we should continue to focus on the degree of variation in the model evaluation results, rather than only considering the detection performance of the model. Therefore, we choose the standard deviation (SD) to evaluate the stability of the model detection results. We can observe that compared with other methods, the SD value of the IS-MLDA reaches the minimum value of 0.99, which indicates that the detection results of our method achieve the minimum degree of fluctuation on unseen domains. Due to the phenomenon of overfitting, other object detection algorithms have achieved high accuracy in the mixed domain D, but the accuracy in the unseen domain has dropped significantly. This clearly proves that our proposed method can effectively improve the generalization and detection

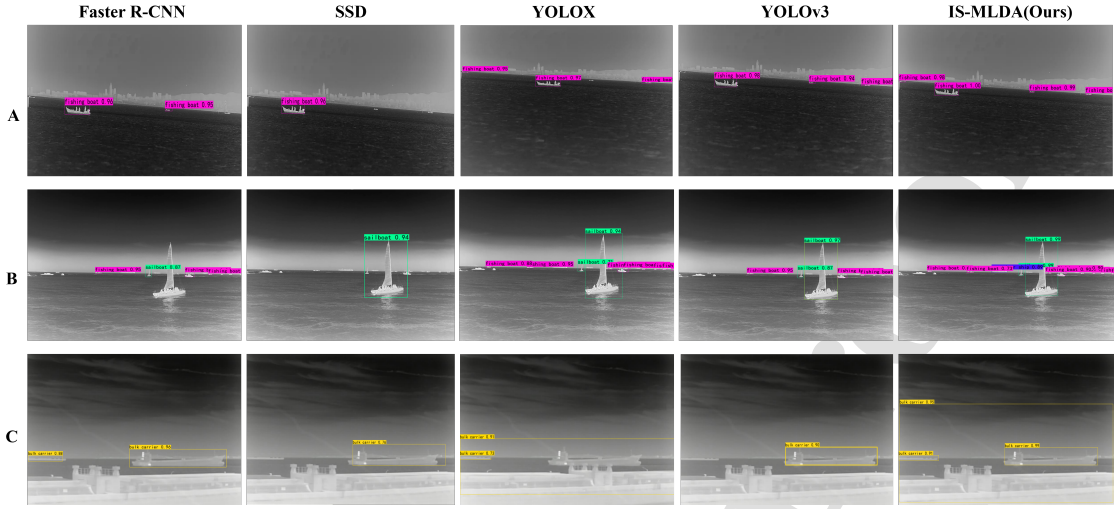


Fig. 7. Detection results of different methods on different target domains.

performance of the model. Fig. 7 shows the detection results of different object detection methods on different target domains, and the detection results of the proposed method are the last column. The detection results of other methods for infrared images in different target domains have different degrees of missed detection, and it is almost difficult to detect small-scale targets in the sea. In contrast, IS-MLDA has better detection performance and stronger generalization performance for images in different target domains. It detects almost all ship targets and has higher detection accuracy for targets of different scales.

To further validate the superiority of IS-MLDA in addressing domain shift, we chose the SOTA domain adaptative object detection (DAOD) methods as our comparison methods, including DAF[52], SWDA[53], S-DAYOLO[54], and MLDG+YOLOv3[34]. From the Table 4, it can be seen that the proposed method achieves the highest detection accuracy on datasets AC $\rightarrow$ B and AB $\rightarrow$ C. Moreover, compared to the currently advanced DAOD methods, IS-MLDA improves the detection accuracy by at least 0.21%, achieving high-precision infrared ship target detection. Specifically, SD reaches a minimum value of 0.99, indicating that the proposed method significantly overcomes the impact of domain shift, effectively enhancing the model's generalization ability. More importantly, there is no target domain data in our training data. Even so, our model still has excellent detection performance on the target domain. According to our proposed learning approach, when a new infrared detector is added to the scene, it is not necessary to train the model with new data again, thus avoiding the complicated manual annotation and the process of needing a large amount of data to drive the model.

4.3.3. Verification of multi-scale object detection performance of IS-MLDA

In order to verify the detection ability of the model for multi-scale ship targets, we take the three domains as the source domains, divide the test sets from the three source domain data respectively, and use them as the mixed domain, and then verify the effectiveness of MFA-s on the mixed domain. We used AP with IOU threshold of 0.5 as the evaluation metric.

As shown in Table 5, even without MFA-s, IS-MLDA outperforms other methods in detection accuracy on the mixed domain. By adding the MFA-s module, the detection performance of IS-MLDA is further enhanced, and the detection accuracy of objects of different scales is improved. Compared with the original IS-MLDA, the detection accuracy with MFA-s is improved by 8.14%, 1.92%, and 2.06% on small, medium, and large-scale targets, respectively.

Next we check if MFA-s can still effectively fuse multi-scale features of ship targets without meta-learning and domain discriminator. We add the MFA-s module to the baseline model YOLOv3 to independently verify the effectiveness of the MFA-s module. As shown in Table 6, without the effect of meta-learning and domain discriminator, our proposed MFA-s can effectively fuse the multi-scale features of the ship target, enhance the multi-scale object

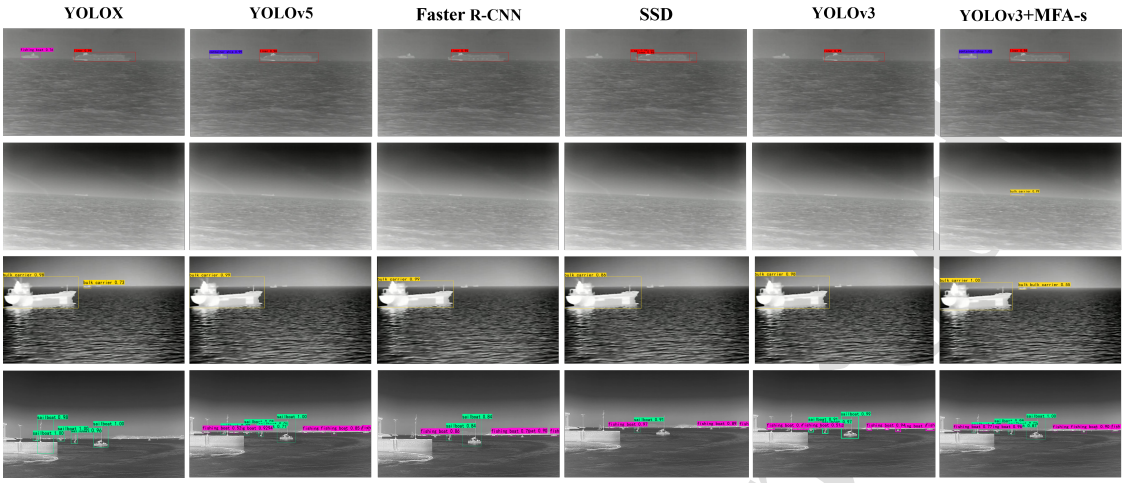


Fig. 8. Visualization of the detection results of different methods.

Table 3: Detection results of infrared images by different object detection algorithms. The data of A, B and C domains are respectively from the data collected by different infrared detectors, domain D is a mixed domain of partial data from the source domain. They have the same detection task but different data distribution. AVG and SD represent the mean and standard deviation of each method respectively, in order to verify the generalization ability of the model on unseen domains, we only calculated AVG and SD on A, B and C. Time indicates the time for the model to detect each image.

Unseen domain	A (%)	B (%)	C (%)	D (%)(mixed-domain)	AVG	SD	Time (ms)
Faster R-CNN	75.43	79.6	78.03	86.19	77.98	2.11	0.42
SSD	64.97	60.71	62.23	68.42	62.64	2.15	0.016
YOLOX	**90.78**	86.94	91.45	95.27	89.72	2.43	0.030
YOLOv3	79.96	71.45	76.89	84.75	76.10	4.31	**0.012**
IS-MLDA(w/o MFA-s)	92.84	**91.03**	**93.36**	94.09	**92.41**	**0.99**	0.013

detection accuracy of the baseline model. Compared with other methods, our method achieves the best detection accuracy on large-scale and medium-scale objects, and the detection accuracy of small-scale objects is close to that of YOLOv5 model, which has advantages in small-scale object detection. Compared with the original YOLOv3 model, our detection accuracy is improved by 2%. At the same time, the AP values of different scales are also improved, especially the detection accuracy of small-scale objects is increased by 4%. This proves once again that the MFA-s module can effectively fuse local features and global features, increase the receptive field of the model, and enhance the feature expression ability of the model. With the improvement of detection accuracy, there is no obvious loss of detection time, which can meet the real-time requirements and provide corresponding technical support for multi-scale target detection in the field of intelligent ship visual perception.

Fig. 8 shows the detection results of different object detection algorithms. We can obviously see that object detection algorithms such as YOLOX and YOLOv5 have different degrees of missed detection for small-scale object and infrared object with low contrast. The YOLOv3 model with MFA-s module detects almost all ship targets, balances the detection accuracy between ship targets of different scales, and ensures the safety of ship navigation at night.

Table 4: Comparison Results with Different DAOD and DG Methods

Unseen domain	A (%)	B (%)	C (%)	D (%)(mixed-domain)	AVG	SD	Time (ms)
DAF	84.03	81.50	85.65	86.34	83.72	2.09	0.462
SWDA	88.97	86.10	89.17	88.08	62.64	1.72	0.457
S-DAYOLO	**92.87**	90.73	93.01	93.74	92.20	1.27	**0.012**
MLDF+YOLOv3	89.06	87.13	89.78	92.53	89.30	1.12	0.013
IS-MLDA(w/o MFA-s)	92.84	**91.03**	**93.36**	**94.49**	**92.41**	**0.99**	0.013

Table 5: Detection results of multi-scale targets by IS-MLDA. The training data comes from partial infrared images in domains A, B and C, and the test data is the mixed test set of domains A, B and C. AP_large, AP_medium, and AP_small is the detection accuracy of large-scale, medium-scale, and small-scale targets.

Method	mAP0.5	AP_large (%)	AP_medium (%)	AP_small (%)	Time (ms)
Faster R-CNN	85.92	82.18	76.9	17.01	0.42
SSD	69.33	69.48	56.41	11.47	0.016
YOLOX	84.90	85.39	77.94	18.53	0.030
YOLOv5	85.76	84.13	78.73	19.67	0.015
YOLOv3	84.28	82.07	76.89	16.35	**0.012**
IS-MLDA(w/o MFA-s)	96.13	94.49	93.35	47.57	0.013
IS-MLDA	**97.81**	**96.55**	**95.27**	**55.71**	0.027

Table 6: Detection results of MFA-s on baseline model. The data used in the training and testing process are all infrared image data.

Method	mAP0.5	AP_large (%)	AP_middle (%)	AP_small (%)	Time (ms)
Faster R-CNN	85.92	82.18	76.9	17.01	0.39
SSD	69.33	69.48	56.41	11.47	0.02
YOLOX	84.90	**85.39**	77.94	18.53	0.030
YOLOv5	85.76	84.13	78.73	19.67	0.015
YOLOv3	84.28	82.07	76.89	16.35	**0.012**
Ours (YOLOv3+MFA-s)	**85.88**	85.28	**78.92**	**20.13**	0.027

4.4. Ablation study

In order to better verify the contribution of each component in our method to the object detection task of infrared images, we conduct the ablation study as follows. The result obtained using meta-learning on the baseline model is denoted by "+ML"; The result obtained by combining meta-learning and domain discriminator is denoted by "+DD"; The result with meta-learning, domain discriminator and MFA-s is denoted by "+MFA-s". We perform ablation study based on the setup in Section 4.3.1, as shown in Table 7. Each method component effectively improves the detection performance of the model. Specifically, the meta-learning algorithm improved the detection accuracy of IS-MLDA on unseen domains by 13.20%. On the basis of meta-learning, we combine the domain discriminator to optimize the domain-invariant features extracted by the model to avoid the overfitting of the model on the source domain. This proves that IS-MLDA performs well in overcoming the influence of domain shift on the deep learning model, can

Table 7: The results of ablative experiment. The data of A, B and C domains are respectively from the data collected by different infrared detectors, domain D is a mixed domain of partial data from the source domain. They have the same detection task but different data distribution. AVG and SD represent the mean and standard deviation of each method respectively, in order to verify the generalization ability of the model on unseen domains, we only calculated AVG and SD on A, B and C. Params is the number of parameters of the model.Time indicates the time for the model to detect each image.

Unseen domain	A (%)	B (%)	C (%)	D (%) (mixed domain)	AVG	SD	Params (M)	Time (s)
YOLOv3	79.96	71.45	76.89	84.75	76.1	4.31	**61.5**	**0.012**
IS-MLDA								
+ML	89.06	87.13	89.78	92.53	89.3	1.12	61.5	0.013
+DD	92.84	91.03	93.36	94.09	92.41	0.99	61.5	0.013
+MFA-s	**94.26**	**93.05**	**94.88**	**95.92**	**94.06**	**0.93**	76.3	0.027

effectively learn the generalization features, and has strong generalization and detection ability. Finally, we added MFA-s to the model. By enhancing the multi-scale features and receptive field of the model, the detection accuracy performance of the model was improved by 1.65%. In addition, the MFA-s module is added to enhance the feature robustness of different scale objects in the training process of meta-learning and domain adaptation, so as to extract the domain independent features of multi-scale objects.

In terms of model parameters, our baseline model YOLOv3 has 61.5M parameters, and the number of our domain discriminator and MFA-s parameters is 2.3M and 13M, respectively. The meta-learning algorithm and domain discriminator in IS-MLDA are only used in the training phase of the model, which means that our method can effectively improve the detection and generalization ability of the model for multi-scale infrared ship targets without increasing a large number of model parameters. In addition, the detection time of the model is close to the original YOLOv3, which meets the engineering application.

To visually illustrate the outcomes after domain alignment, following the extraction of domain-invariant features through meta-learning and domain discriminator, t-SNE was employed for visualizing domain features, as shown in Fig. 6(b). From the figure, it is evident that features from different domains overlap, indicating that IS-MLDA can effectively extract common feature representations across diverse domains. In the feature space, it maps features from different domains to similar distributions, thereby enhancing the model's robustness to variations between different domains.

5. CONCLUSION

In this paper, we proposed a novel infrared ship object detection method IS-MLDA that leverages meta-learning and domain adaptation to tackle the domain shift problem. IS-MLDA consists of three key components: coordinated gradient optimization, domain discriminator, and the MFA-s module. Coordinated gradient optimization on both meta-train and meta-test sets during training was designed to learn domain-invariant representations for infrared ship targets. It can minimize the differences between source and target domains in the feature space, and enhance model generalization capabilities. Domain discriminator was designed to mitigate overfitting on the source domain. It can provide global supervision information during training and effectively filters out domain-specific information. This will help model generalize well to unseen domains. The MFA-s module was introduced to enhance the model receptive field and effectively extract and fuse multi-scale features of ships. By combining local and global features of different scales, the model's feature expression ability was significantly improved, leading to better detection performance for multi-scale targets.

Experimental results demonstrated that IS-MLDA can effectively solve the domain shift problem caused by inconsistent data distribution between different domains. It achieves significantly improved infrared ship target detection accuracy with high efficiency. Specifically, the average mAP on the target domain was increased by 18%, SD reached a minimum of 0.93. The proposed method showed excellent generalization ability and can greatly improve navigation safe of intelligent ships, especially during nighttime. In future we will explore the integration of complementary

information from visible and infrared images. It will further improve the precision of ship target detection and enable accurate object recognition under all-weather conditions.

CRediT authorship contribution statement

Hui Feng: Conceptualization, Methodology, Writing - review & editing. **Wei Tang**: Conceptualization, Methodology, Experiments, Writing - original draft. **Haixiang Xu**: Conceptualization, Writing - review & editing. **Chengxin Jiang**: Conceptualization. **Sam Shuzhi Ge**: Conceptualization, Writing - review & editing. **Jianhua He**: Conceptualization, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors appreciate the constructive suggestions from reviewers and the Handling Editor. This work is supported by the National Natural Science Foundation of China under Grant No. 52371374, 51979210. This work is partly funded by EPSRC with RC Grant No. EP/Y027787/1,UKRI under grant No. EP/Y028317/1, the Horizon Europe MSCA programme under grant agreement No. 101086228. For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript (AAM) version arising from this submission.

Data availability

Data will be made available on request. For specific access to the data, please contact our corresponding author via email at qukaiyang-whut@163.com.

References

[1] C. Chen, X.-Q. Chen, F. Ma, X.-J. Zeng, and J. Wang, "A knowledge-free path planning approach for smart ships based on reinforcement learning," *Ocean Engineering*, vol. 189, p. 106299, 2019.

[2] J.-J. Liu, Q. Hou, M.-M. Cheng, J. Feng, and J. Jiang, "A simple pooling-based design for real-time salient object detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 3917–3926, 2019.

[3] H. Feng, J. Guo, H. Xu, and S. S. Ge, "Sharpgan: dynamic scene deblurring method for smart ship based on receptive field block and generative adversarial networks," *Sensors*, vol. 21, no. 11, p. 3641, 2021.

[4] J. Guo, H. Feng, H. Xu, W. Yu, and S. shuzhi Ge, "D3-net: integrated multi-task convolutional neural network for water surface deblurring, dehazing and object detection," *Engineering Applications of Artificial Intelligence*, vol. 117, p. 105558, 2023.

[5] X. Bai, M. Liu, T. Wang, Z. Chen, P. Wang, and Y. Zhang, "Feature based fuzzy inference system for segmentation of low-contrast infrared ship images," *Applied Soft Computing*, vol. 46, pp. 128–142, 2016.

[6] T. Chen, G. Fu, S. Li, and Y. Li, "Typical target detection for infrared homing guidance based on yolo v3," *Laser Optoelectronics Progress*, vol. 56, no. 16, p. 161502, 2019.

[7] B. Zhao, C. Wang, Q. Fu, and Z. Han, "A novel pattern for infrared small target detection with generative adversarial network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 5, pp. 4481–4492, 2020.

[8] D. Tuia, C. Persello, and L. Bruzzone, "Recent advances in domain adaptation for the classification of remote sensing data," *arXiv preprint arXiv:2104.07778*, 2021.

[9] J. M. Mooney and F. D. Shepherd, "Characterizing ir fpa nonuniformity and ir camera spatial noise," *Infrared Physics Technology*, vol. 37, no. 5, pp. 595–606, 1996.

[10] X. Chen, S. Wang, M. Long, and J. Wang, "Transferability vs. discriminability: Batch spectral penalization for adversarial domain adaptation," in *International conference on machine learning*, pp. 1081–1090, PMLR, 2019.

[11] R. Gong, W. Li, Y. Chen, and L. V. Gool, "Dlow: Domain flow for adaptation and generalization," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2477–2486, 2019.

[12] K. Saito, Y. Ushiku, T. Harada, and K. Saenko, "Strong-weak distribution alignment for adaptive object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6956–6965, 2018.

[13] W. Li, X. Liu, and Y. Yuan, "Sigma: Semantic-complete graph matching for domain adaptive object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5291–5300, 2022.

[14] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *Journal of Machine Learning Research*, vol. 17, no. 1, pp. 2096–2030, 2016.

[15] K. Zhou, Y. Yang, T. Hospedales, and T. Xiang, "Deep domain-adversarial image generation for domain generalisation," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, pp. 13025–13032, 2020.

[16] M. S. Long, H. Zhu, J. M. Wang, and M. I. Jordan, "Unsupervised domain adaptation with residual transfer networks," in *30th Conference on Neural Information Processing Systems (NIPS)*, vol. 29 of *Advances in Neural Information Processing Systems*, (LA JOLLA), Neural Information Processing Systems (Nips), 2016.
[17] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7167–7176, 2017.
[18] B. Zhang, Y. Tan, H. Wang, Z. Zhang, X. Zhou, J. Wu, Y. Mi, H. Huang, and W. Wang, "Lsrml: A latent space regularization based meta-learning framework for mr image segmentation," *Pattern Recognition*, vol. 130, p. 108821, 2022.
[19] X. Zhang, P. Cui, R. Xu, L. Zhou, Y. He, and Z. Shen, "Deep stable learning for out-of-distribution generalization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5372–5382, 2021.
[20] P. Pandey, M. Raman, S. Varambally, and P. AP, "Domain generalization via inference-time label-preserving target projections," *arXiv preprint arXiv:2103.01134*, 2021.
[21] F. S. Marvasti, M. R. Mosavi, and M. Nasiri, "Flying small target detection in ir images based on adaptive toggle operator," *IET Computer Vision*, vol. 12, no. 4, pp. 527–534, 2018.
[22] X. Y. Wang, Z. M. Peng, P. Zhang, and Y. M. He, "Infrared small target detection via nonnegativity-constrained variational mode decomposition," *Ieee Geoscience and Remote Sensing Letters*, vol. 14, no. 10, pp. 1700–1704, 2017.
[23] B. Wang, E. Benli, Y. Motai, L. Dong, and W. Xu, "Robust detection of infrared maritime targets for autonomous navigation," *IEEE Transactions on Intelligent Vehicles*, vol. 5, no. 4, pp. 635–648, 2020.
[24] B. Bosquet, M. Mucientes, and V. M. Brea, "Stdnet: Exploiting high resolution feature maps for small object detection," *Engineering Applications of Artificial Intelligence*, vol. 91, p. 103615, 2020.
[25] H. F. Wang, J. Y. Jiang, Q. Zhao, H. Li, K. Yan, Y. Yang, S. L. Li, Y. G. Zhang, L. L. Qiao, C. L. Fu, H. Yin, Y. Hu, and H. B. Yu, "Progressive structure network-based multiscale feature fusion for object detection in real-time application," *Engineering Applications of Artificial Intelligence*, vol. 106, 2021.
[26] R. Miao, H. Jiang, and F. Tian, "Robust ship detection in infrared images through multiscale feature extraction and lightweight cnn," *Sensors*, vol. 22, no. 3, p. 1226, 2022.
[27] Z. Song, J. Yang, D. Zhang, S. Wang, and Z. Li, "Semi-supervised dim and small infrared ship detection network based on haar wavelet," *IEEE Access*, vol. PP, no. 99, pp. 1–1, 2021.
[28] S. Du, P. Zhang, B. Zhang, and H. Xu, "Weak and occluded vehicle detection in complex infrared environment based on improved yolov4," *IEEE Access*, vol. PP, no. 99, pp. 1–1, 2021.
[29] H. Zhou, F. Jiang, and H. Lu, "Ssda-yolo: Semi-supervised domain adaptive yolo for cross-domain object detection," *Computer Vision and Image Understanding*, vol. 229, p. 103649, 2023.
[30] S. Zhao, Y. Luo, T. Zhang, W. Guo, and Z. Zhang, "A domain specific knowledge extraction transformer method for multisource satellite-borne sar images ship detection," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 198, pp. 16–29, 2023.
[31] S. Shankar, V. Piratla, S. Chakrabarti, S. Chaudhuri, P. Jyothi, and S. Sarawagi, "Generalizing across domains via cross-gradient training," *arXiv preprint arXiv:1804.10745*, 2018.
[32] F. M. Carlucci, A. D'Innocente, S. Bucci, B. Caputo, and T. Tommasi, "Domain generalization by solving jigsaw puzzles," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2229–2238, 2019.
[33] P. Wu, H. Huang, H. Qian, S. Su, B. Sun, and Z. Zuo, "Srcanet: Stacked residual coordinate attention network for infrared ship detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2022.
[34] D. Li, Y. Yang, Y.-Z. Song, and T. Hospedales, "Learning to generalize: Meta-learning for domain generalization," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, 2018.
[35] H. F. Xia, H. D. Zhao, Z. M. Ding, and Ieee, "Adaptive adversarial network for source-free domain adaptation," in *18th IEEE/CVF International Conference on Computer Vision (ICCV)*, (NEW YORK), pp. 8990–8999, Ieee, 2021.
[36] B. Sun and K. Saenko, "Deep coral: Correlation alignment for deep domain adaptation," in *Computer Vision–ECCV 2016 Workshops: Amsterdam*, Computer Vision – ECCV 2016 Workshops, pp. 443–450, Springer International Publishing, 2016.
[37] L. Zhang and Z. Peng, "Infrared small target detection based on partial sum of the tensor nuclear norm," *Remote Sensing*, vol. 11, no. 4, p. 382, 2019.
[38] Y. Lu, D. Li, W. Wang, Z. Lai, J. Zhou, and X. Li, "Discriminative invariant alignment for unsupervised domain adaptation," *IEEE Transactions on Multimedia*, vol. 24, pp. 1871–1882, 2022.
[39] Y. Lu, Q. Zhu, B. Zhang, Z. Lai, and X. Li, "Weighted correlation embedding learning for domain adaptation," *IEEE Transactions on Image Processing*, vol. 31, pp. 5303–5316, 2022.
[40] Y. Lu, W. K. Wong, B. Zeng, Z. Lai, and X. Li, "Guided discrimination and correlation subspace learning for domain adaptation," *IEEE Transactions on Image Processing*, vol. 32, pp. 2017–2032, 2023.
[41] I. Albuquerque, J. Monteiro, M. Darvishi, T. H. Falk, and I. Mitliagkas, "Generalizing to unseen domains via distribution matching," *arXiv preprint arXiv:1911.00804*, 2019.
[42] X. Jin, C. Lan, W. Zeng, and Z. Chen, "Feature alignment and restoration for domain generalization and adaptation," *arXiv preprint arXiv:2006.12009*, 2020.
[43] Y. Li, Y. Yang, W. Zhou, and T. Hospedales, "Feature-critic networks for heterogeneous domain generalization," in *International Conference on Machine Learning*, pp. 3915–3924, PMLR, 2019.
[44] Y. Balaji, S. Sankaranarayanan, and R. Chellappa, "Metareg: towards domain generalization using meta-regularization," in *Neural Information Processing Systems*, 2018.
[45] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," *JMLR.org*, 2014.
[46] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "Yolox: Exceeding yolo series in 2021," *arXiv e-prints*, 2021.
[47] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788, 2016.
[48] T. Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," *IEEE Computer*

Society, 2017.

[49] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. M. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," in *33rd Conference on Neural Information Processing Systems (NeurIPS)*, vol. 32 of *Advances in Neural Information Processing Systems*, (LA JOLLA), Neural Information Processing Systems (Nips), 2019.

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

[51] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, pp. 21–37, Springer, 2016.

[52] Y. Chen, W. Li, C. Sakaridis, D. Dai, and L. Van Gool, "Domain adaptive faster r-cnn for object detection in the wild," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3339–3348, 2018.

[53] K. Saito, Y. Ushiku, T. Harada, and K. Saenko, "Strong-weak distribution alignment for adaptive object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6956–6965, 2019.

[54] G. Li, Z. Ji, X. Qu, R. Zhou, and D. Cao, "Cross-domain object detection for autonomous driving: A stepwise domain adaptative yolo approach," *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 3, pp. 603–615, 2022.

Highlights:

(1) The proposed IS-MLDG facilitates end-to-end detection of infrared ship targets without the need for image preprocessing, while effectively extracting domain-invariant features to enhance the model's generalization ability and elevate detection performance in previously unseen domains.

(2) We pioneer the application of gradient-based meta-learning algorithms and domain adaptation methods in infrared ship object detection, demonstrating the efficacy of combining domain generalization and adaptation to address the domain shift challenge.

(3) A multi-scale feature fusion module is designed to aggregate local and global features, which enhanced the model's feature representation ability.

(4) We develop new datasets for ship detection and domain adaptation, simulating the domain shift phenomenon using infrared ship images from different infrared sensors in real-world scenarios.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.