

An Imbalanced Discriminant Alignment Approach for Domain Adaptive SAR Ship Detection

Bin Pan^{ID}, *Member, IEEE*, Zhehao Xu, Tianyang Shi^{ID}, Tao Li^{ID}, and Zhenwei Shi^{ID}, *Member, IEEE*

Abstract—Synthetic aperture radar (SAR) imaging has round-the-clock data acquisition capability regardless of light and climate constraints, so it has been widely used for ship detection. However, SAR images usually suffer lower imaging quality, which may result in indistinct contours and non-negligible noise. Therefore, the manual labeling for SAR images is expensive, leading to a lack of training data in the task of ship detection. In this article, we propose a route by utilizing domain adaptive methods to transfer information from labeled visible images (source domain) to unlabeled SAR images (target domain) for ship detection. To address the distribution mismatch between domains, we develop a novel imbalanced discriminant alignment (IDA) approach to improve the discriminant ability of the network and prevent negative migration. The core of the IDA approach is applying a new loss function called imbalanced prediction consistency (IPC) loss to describe the domain classifier consistency, and we further provide theoretical analysis for the effectiveness of the IPC loss. IDA ensures consistency at the image level and instance level, and focuses on the consistency of the source domain to enhance the feature extraction capability of the adversarial network. The theoretical discussion has proven that a necessary and sufficient condition for convergence of the IPC loss is that the two discriminant probabilities converge to 0 at the discriminant distance we define. Experimental results have indicated the advantage of IDA when compared with other domain adaptation SAR ship detection methods.

Index Terms—Adversarial learning, discriminant alignment loss, domain adaptation, synthetic aperture radar (SAR) ship detection.

I. INTRODUCTION

SYNTHETIC aperture radar (SAR) applies successive pulses of microwave and the echo of each pulse imaging, which can avoid weather-caused signal attenuation and has day and night data collection capability [1]. Thus, SAR has been widely used in remote sensing target detection tasks [2], [3]. Among many target detection applications, ship

detection [4], [5], [6] is one of the meaningful tasks which aims at finding ships from the sea and the harbor. However, the unique imaging mode of SAR brings in some shortcomings for ship detection. As shown in Fig. 1, compared with the images generated by visible light sensors, SAR images have a lower signal-to-noise ratio and spatial resolution, and severe image noise contamination, which leads to difficulty for ship detection.

Conventional SAR ship detection mainly included two categories [7]: 1) approaches based on constant false alarm rate (CFAR) and 2) approaches based on machine learning and deep learning. The emergence of the Biparameter CFAR detector brought SAR image target detection to light [8]. Approaches based on CFAR leveraged the radar scattering cross-sectional area information to select the bright contrast part as the candidate region. The following methods were based on CFAR to extract regional accuracy and improve computational efficiency [9], [10], [11]. Furthermore, some approaches combined CFAR into various features such as geometric structure, extended fractal, and wavelet coefficients to fuse to detect the target [12], [13], [14]. In recent years, with the advancement of computing hardware and the discovery of the feature expression capability of convolution, techniques based on machine learning and deep learning have become one of the promising choices in SAR target detection. Classical machine learning methods such as support vector machines and Adaboost were used to screen the whole image region to get the potential region and combined with the signal processing approach to refine the location [15], [16], [17], [18], [19]. For the characteristics of SAR images, deep learning techniques were used for target detection by classical convolutional networks [20], [21] along with heterogeneous innovations, such as dense connectivity [2], [22], [23], feature reconstruction [24], [25], and hierarchical alignment [26], [27].

However, obtaining sufficient annotated samples for supervised training remains challenging. SAR images are more difficult to annotate than visible remote sensing images because of their unique appearance and expensive collection costs. SAR images have a speckled appearance caused by the radar scattering off different surfaces and have a diverse polarization compared to optical images, which makes it more challenging to annotate accurately [28].

Domain adaptation is a technique that may contribute to solving the problem of insufficient annotated samples [29]. The goal of domain adaptation is to train a model on one dataset and ensure good accuracy on another dataset whose distribution is significantly different from the former.

Manuscript received 1 March 2023; revised 27 June 2023; accepted 31 July 2023. Date of publication 9 August 2023; date of current version 23 August 2023. This work was supported in part by the National Key Research and Development Program of China under Grant 2022ZD0160401 and Grant 2022YFA1003803; in part by the National Natural Science Foundation of China under Grant 62001251, Grant 62001252, Grant 62125102, and Grant 62272248; and in part by the Beijing-Tianjin-Hebei Basic Research Cooperation Project under Grant F2021203109. (Corresponding author: Tao Li.)

Bin Pan and Zhehao Xu are with the School of Statistics and Data Science, KLMDASR, LEBPS, and LPMC, Nankai University, Tianjin 300071, China (e-mail: panbin@nankai.edu.cn; xuzhehao@mail.nankai.edu.cn).

Tianyang Shi and Zhenwei Shi are with the Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, China (e-mail: shitianyang@buaa.edu.cn; shizhenwei@buaa.edu.cn).

Tao Li is with the College of Computer Science, Nankai University, Tianjin 300071, China (e-mail: litao@nankai.edu.cn).

Digital Object Identifier 10.1109/TGRS.2023.3303507

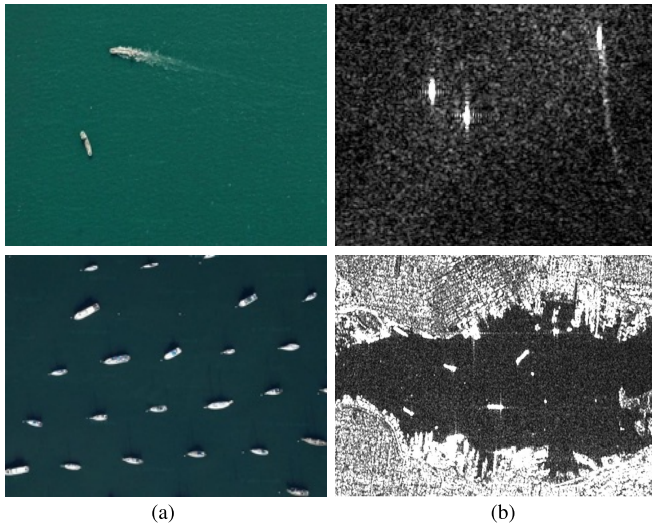


Fig. 1. Illustration for discrepancy of (a) visible and (b) SAR images. Visible images are distinct and elementary to be annotated, but SAR images are laborious for manual annotation due to their special imaging mode.

Generally, we refer to the cluster of data distribution utilized for training as the source domain and another data distribution cluster for testing as the target domain [30]. Recently, several studies have employed domain adaptation methods for SAR image target detection. For instance, Shi et al. [31] integrated multiple discriminators into the pioneering work of domain-adapted target detection proposed by Chen et al. [32]. Zhang et al. [26] switched frequency images generated by fast Fourier transform for cross-domain adaptation. Song et al. [33] first leveraged CycleGAN [34] to transfer the style of the source domain and then used the popular domain adaptation algorithm for target detection. However, the significant difference in the distribution of SAR and visible images can result in negative migration or training instability during domain adaptation. Although the two-stage style migration methods can address these issues, they may bring in cumulative error when compared with end-to-end networks.

In this article, we propose an end-to-end domain adaptation-based ship detection network, imbalanced discriminant alignment (IDA), to address the negative migration challenge which results from the significant distribution discrepancy between the two domains. IDA focuses on the imbalanced alignment of discriminant probabilities from the source and target domains by utilizing imbalanced prediction consistency (IPC) loss, and is theoretically proven to be effective in overcoming the negative migration phenomenon. Two distributions with lower correlation have a greater probability of encountering negative migration because the network is more challenging to come up with a valid representation of the features. Our motivation is that adversarial learning-based domain adaptation methods [26], [31], [33] usually consider that the information in source and target domains has equal influence. However, incorrect information on the target domain without labels may make it challenging for the network to learn a valid feature representation. Therefore, IDA is leveraged to imbalance in adversarial learning, enabling the network to make better use of valid information from the

source domain and less use of incorrect information from the target domain. Moreover, we provide theoretical analysis for the effectiveness of IPC loss, where the new loss presents weaker discriminant consistency in the target domain and stronger in the source domain. In summary, the contributions of this article are listed below.

- 1) We propose a new domain adaptation-based ship detection method for SAR images, where the IDA approach is developed to deal with the negative migration problem when the distribution of visible and SAR images differs sharply.
- 2) We have theoretically analyzed and proven the imbalanced alignment of IPC loss in the heterogeneous domains. The IPC loss helps IDA implement strict alignment on the source domain while soft alignment on the target domain.

II. METHODOLOGY

In this section, we briefly introduce the preliminaries of our proposed framework, and then we provide a detailed explanation of the IPC loss used in the IDA approach. Ultimately, we present the credibility of our proposed method from the theoretical perspective.

Following the common terminology in unsupervised domain adaptation [30], we are given the labeled $\mathcal{D}_S = \{(x_i, y_i)\}_{i=1}^{n_s}$ as the source domain for training and unlabeled $\mathcal{D}_T = \{x_i\}_{i=1}^{n_t}$ as the target domain for testing. We aim to extract representative invariant features from the source and target domains in the shared space and to avoid negative migration, which we achieve using the adversarial network with the IDA approach.

A. Preliminaries

The architecture of our proposed method is illustrated in Fig. 2. Our work follows the main idea of domain adaptive Faster R-CNN (DA model) [32], which serves as the baseline for domain adaptation using adversarial learning. The DA model consists of two major components: image-level adaptation D_{img} and instance-level adaptation D_{ins} . Our proposed method integrates a domain classifier consistency module into these two components to ensure discriminant consistency through the use of the IDA approach.

The input image from the source or target domain is processed using a shared Faster R-CNN backbone [35], including convolutional layers for image-level feature extraction. Instance-level features are obtained through the region proposal network (RPN) and a region-of-interest (ROI) module in the backbone. To enable adversarial learning, we employ a strategy where the image-level and instance-level feature extractors, along with discriminators D_{img} and D_{ins} , compete against each other. This entails optimizing the feature extractors to maximize the loss while the discriminators aim to minimize the loss. The gradient reverse layer (GRL) [36] automatically reverses the gradient during backpropagation, allowing image-level and instance-level alignment modules to contend with the backbone using an inverse optimization direction. The backbone can extract shared features from both domains through a max-min process, thereby enhancing the

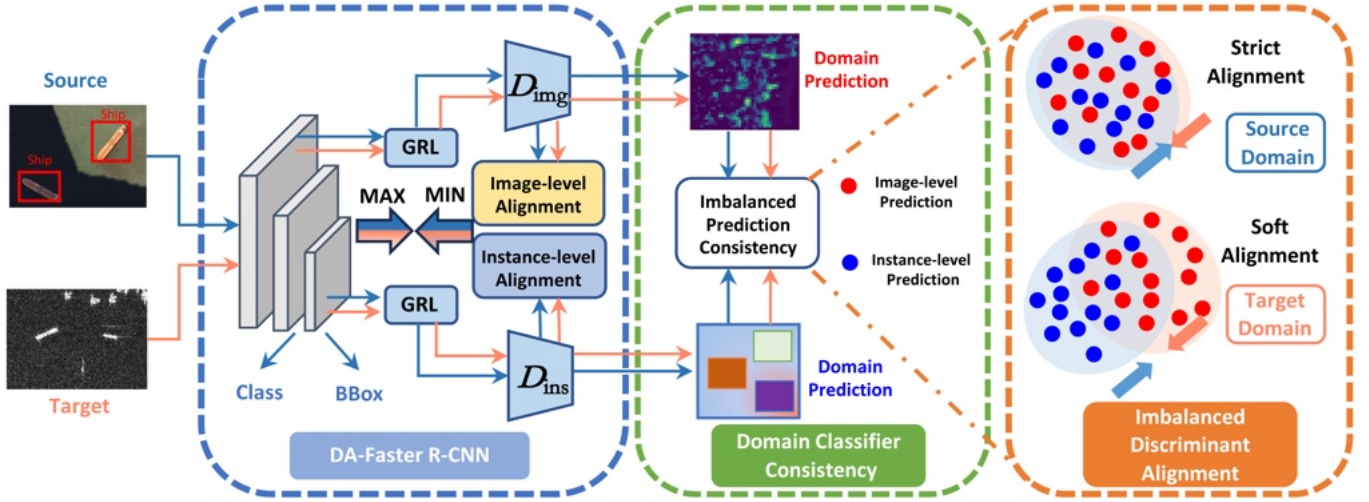


Fig. 2. Illustration of unsupervised domain adaptation for SAR images target detection based on DA Faster R-CNN. D_{img} and D_{ins} denote image-level and instance-level feature extractors behind GRL layers to achieve adversarial learning. A domain classifiers consistency module is incorporated into these two domain predictions to overcome the distribution discrepancy challenge using the IDA approach, which makes strict alignment on the source domain while soft alignment on the target domain. In the case of significant differences in domain distribution, this alignment with different strengths improves the robustness of the feature extraction networks.

accuracy of the detector. During adversarial learning, D_{img} and D_{ins} individually estimate the domain of the input features, and their judgments should align for effective optimization. To address inconsistencies, discriminant consistency modules are added after domain adaptation. It is important to note that we utilize the shared backbone detection header for classification and regression (Class and BBox in Fig. 2). In the training process, we solely optimize the classification and regression for the source domain as labels for the target domain are unavailable.

We briefly introduce the two core modules in the DA model, image-level alignment and instance-level alignment, which are significant in our proposed network.

1) *Image-Level Alignment*: For the global features extracted from the backbone, we utilize a domain discriminator D_{img} to output the domain prediction map. The loss of image-level alignment can be formatted as

$$L_{\text{img}} = - \sum_{i,u,v} [D_i \log p_i^{u,v} + (1 - D_i) \log(1 - p_i^{u,v})]. \quad (1)$$

In (1), D_i refers to the domain label of the i th image, where $D_i = 0$ for the source domain and $D_i = 1$ for the target domain. The probability prediction, denoted as $p_i^{u,v}$, represents the prediction of the pixel located at coordinates (u, v) on the output map. $p_i^{u,v}$ is output by the Softmax function, thus $p_i^{u,v} \in (0, 1)$. Both the prediction map and the input feature have matching width and height elements.

2) *Instance-Level Alignment*: We also utilize a domain discriminator D_{ins} to output the domain prediction for the instance features extracted from RPN and ROI. Nonetheless, D_{ins} only judges each instance extracted from the network. The loss of instance-level alignment can be summarized as

$$L_{\text{ins}} = - \sum_{i,j} [D_i \log p_{i,j} + (1 - D_i) \log(1 - p_{i,j})] \quad (2)$$

where $p_{i,j}$ denotes the probability prediction that the j th proposal region in the i th image. $p_{i,j}$ is also output by the Softmax function, thus $p_{i,j} \in (0, 1)$.

The DA model is based on Faster R-CNN for domain adaptation improvement. The goals to be optimized for Faster R-CNN are classification (Class) loss L_{cls} and bounding box regression (BBox) loss L_{reg} . The classification loss L_{cls} is in the form of cross-entropy for the target probability prediction, while the regression loss is in the form of smooth ℓ_1 loss for predicting the target offset parameters and the actual offset parameters. The specific form of these losses function is detailed in [35].

B. IPC Loss

The existing theories show that enforcing consistency image-level and instance-level prediction can enhance the cross-domain robustness of the network [32]. To further heighten this consistency, we define a more rigorous way of discriminant consistency measuring.

Definition 1 (Discriminant Distance): Assume p and q are two discriminant probability with $p \in (0, 1)$ and $q \in (0, 1)$, the discriminant distance $D_{\text{dis}}(p, q)$ is defined as

$$D_{\text{dis}}(p, q) \triangleq \max\left(\frac{p}{q}, \frac{q}{p}\right). \quad (3)$$

From the definition of discriminant distance, we can infer that $D_{\text{dis}} \geq 1$. Specially, $D_{\text{dis}}(p, q) = 1$ if and only if $p = q$. And for another discriminant probability $r \in (0, 1)$, if $D_{\text{dis}}(p, q) < D_{\text{dis}}(p, r)$, it means (p, q) have more discriminant consistency than (p, r) . A smaller discriminant distance means that the two discriminant probabilities are more consistent.

In comparison to minimizing the l_1 or l_2 distance to 0, the discriminant distance has more stringent requirements for alignment. In other words, the discriminant distance we define is more general than the l_1 or l_2 distance. In fact, we have the theorem below.

Theorem 1: $D_{\text{dis}}(p, q) \rightarrow 1$ is the sufficient and unnecessary condition for $\|p, q\| \rightarrow 0$. $\|\cdot, \cdot\|$ is l_1 or l_2 distance, and p and q are two continuous variables with $p \in (0, 1)$ and $q \in (0, 1)$.

Proof: Sufficiency: From $D_{\text{dis}}(p, q) \rightarrow 1$ and the definition of discriminant distance, we infer that $p/q \rightarrow 1$. Hence, $(p - q)/q \rightarrow 0$. Due to $q \in (0, 1)$, we have $p - q \rightarrow 0$. Consequently, $\|p, q\| \rightarrow 0$.

Unnecessity: Let $p = 2q$ and $q \rightarrow 0$, we have $\|p, q\| \rightarrow 0$. On the other hand, $D_{\text{dis}}(p, q) \equiv 2 \neq 1$. \square

Additionally, in domain adaptation tasks, the target domain lacks labels, so the data from the source domain typically has a more profound impact on the network. Therefore, we aim to treat the data unevenly: we consider more strict alignment for instance-level and image-level features to enhance the extraction capability of the network for the source domain. On the other hand, for the target domain, we loosen alignment to prevent the network from over-fitting and extracting ineffective features. Therefore, we propose the following IPC loss based on discriminant distance:

$$L_{\text{ipc}} \triangleq \sum_{i,j} \frac{|q_i - p_{i,j}|}{|q_i - p_{i,j}| + \beta \max(q_i, p_{i,j}) + (1 - \beta) \min(q_i, p_{i,j})} \quad (4)$$

where

$$\frac{1}{|I|} \sum_{u,v} p_i^{u,v} = q_i \quad (5)$$

and $|I|$ denotes the number of pixels in the feature map. $\beta \in [0, 1]$ is the hyperparameter used for imbalance. $p_i^{u,v}$ and $p_{i,j}$ are outputs of the Softmax layer in D_{img} and D_{ins} , so $q_i, p_{i,j} \in (0, 1)$. It is worth noting that our bounds are all open intervals in our theorem. This is because our discriminant probabilities are output by the Softmax function. Therefore, the obtained probabilities are open intervals on 0 to 1 rather than closed intervals.

C. IDA Approach

The IDA method enables the learning process of networks that avoids negative migration by limiting the discriminant probability of the two discriminator outputs using IPC loss. First of all, we prove that the optimization objective of the IPC loss function is minimizing the discriminant distance of the instance-level and image-level discriminant probabilities.

Theorem 2: For the IPC loss L_{ipc} defined in (4), A sufficient and necessary condition for $L_{\text{ipc}} \rightarrow 0$ is: for every j th instance in i th image, $D_{\text{dis}}(q_i, p_{i,j}) \rightarrow 1$.

Proof: Sufficiency: assume that $q_i \neq p_{i,j}$. Otherwise, we have reached the optimal state. We rewrite the IPC loss in (4)

$$L_{\text{ipc}} = \sum_{i,j} \frac{|q_i - p_{i,j}|}{|q_i - p_{i,j}| + \beta \max(q_i, p_{i,j}) + (1 - \beta) \min(q_i, p_{i,j})}$$

$$\begin{aligned} &= \sum_{i,j} \left\{ \frac{|q_i - p_{i,j}|}{|q_i - p_{i,j}|} + \frac{\beta}{2} (|q_i + p_{i,j}| + |q_i - p_{i,j}|) + \frac{1 - \beta}{2} (|q_i + p_{i,j}| - |q_i - p_{i,j}|) \right\} \\ &= \sum_{i,j} \frac{\frac{|q_i - p_{i,j}|}{2} + \frac{1 + 2\beta}{2} |q_i - p_{i,j}|}{\frac{|q_i + p_{i,j}|}{2} + \frac{1 + 2\beta}{2} |q_i - p_{i,j}|} \\ &= \sum_{i,j} \frac{2|q_i - p_{i,j}|}{|q_i + p_{i,j}| + (1 + 2\beta)|q_i - p_{i,j}|} \\ &= \sum_{i,j} \frac{2}{\frac{|q_i + p_{i,j}|}{|q_i - p_{i,j}|} + (1 + 2\beta)}. \end{aligned} \quad (6)$$

To facilitate the prove, since both q_i and $p_{i,j}$ are real numbers, we denote $q_i = k_{i,j} p_{i,j}$, $f_{i,j} = (|q_i + p_{i,j}|/|q_i - p_{i,j}|)$. Hence

$$f_{i,j} = \frac{|k_{i,j} p_{i,j} + p_{i,j}|}{|k_{i,j} p_{i,j} - p_{i,j}|} = \frac{|k_{i,j} + 1|}{|k_{i,j} - 1|} = \left| 1 + \frac{2}{k_{i,j} - 1} \right|. \quad (7)$$

From (7), $f_{i,j}$ is the elementary combination of inverse proportional function, we can infer that $f_{i,j} \rightarrow +\infty \iff k_{i,j} \rightarrow 1$. Meanwhile, $D_{\text{dis}}(q_i, p_{i,j}) \rightarrow 1$ is equivalent to $k_{i,j} \rightarrow 1$. Therefore, the denominator in each term of (6) tends to positive infinity when $D_{\text{dis}}(q_i, p_{i,j}) \rightarrow 1$.

Additionally, the number of input images is limited, and the instances in each image are also limited. Hence, the number of terms in (6) is limited. Above all, if $D_{\text{dis}}(q_i, p_{i,j}) \rightarrow 1$, each term in (6) tends to 0, and this limited number of terms summation leads to $L_{\text{ipc}} \rightarrow 0$.

Necessity: from the proof of sufficiency, we know that L_{ipc} has finite terms, and every term in L_{ipc} is non-negative. Therefore, $L_{\text{ipc}} \rightarrow 0$ requires every term in L_{ipc} to tends to 0, which means $k_{i,j} \rightarrow 1$. This is equivalent to $D_{\text{dis}}(q_i, p_{i,j}) \rightarrow 1$. \square

Therefore, we know from Theorem 2 that limiting the discriminator prediction consistency can be realized by optimizing the IPC loss to 0, which is adopted by the IDA approach. Furthermore, from a gradient perspective, we illustrate why IDA approach can provide more stringent requirements than the normal l_2 loss. First, compute the gradient of $|q_i - p_{i,j}|$ with respect to q_i

$$\nabla_{q_i} |q_i - p_{i,j}| = \frac{q_i - p_{i,j}}{|q_i - p_{i,j}|} = \text{sgn}(q_i - p_{i,j}) \quad (8)$$

where

$$\text{sgn}(x) = \begin{cases} 1 & x > 0 \\ y & x = 0 \\ -1 & x < 0 \end{cases} \quad (9)$$

y is a random value in $[-1, 1]$. Therefore, the gradient of L_{ipc} with respect to q_i can be written as

$$\begin{aligned} \nabla_{q_i} L_{\text{ipc}} &= \sum_j \left\{ (2\text{sgn}(q_i - p_{i,j})) (q_i + p_{i,j} + (1 + 2\beta)|q_i - p_{i,j}|) \right\} \end{aligned}$$

$$\begin{aligned}
& -2|q_i - p_{i,j}|(1 + (1 + 2\beta)\text{sgn}(q_i - p_{i,j})) \\
& / (|q_i + p_{i,j}| + (1 + 2\beta)|q_i - p_{i,j}|)^2 \Big\} \\
= & \sum_j \frac{2\text{sgn}(q_i - p_{i,j})(q_i + p_{i,j}) - 2|q_i - p_{i,j}|}{(|q_i + p_{i,j}| + (1 + 2\beta)|q_i - p_{i,j}|)^2}. \quad (10)
\end{aligned}$$

Similarly, the gradient of L_{ipc} with respect to $p_{i,j}$ can be written as

$$\nabla_{p_{i,j}} L_{\text{ipc}} = \frac{2\text{sgn}(q_i - p_{i,j})(q_i + p_{i,j}) - 2|q_i - p_{i,j}|}{(|q_i + p_{i,j}| + (1 + 2\beta)|q_i - p_{i,j}|)^2}. \quad (11)$$

In (10) and (11), the numerator is the primary form of discriminant probability, while the denominator is quadratic. Therefore, the IPC loss has a larger gradient than the l_2 loss with the same discriminant probability. So in the optimization process, our loss provides a closer discriminant distance.

Finally, we explain the unbalanced effect and advantage of our proposed approach. Actually, the unbalance is caused by the discriminant distance. For source domain data, our discriminant probability is closer to 0, while for target domain data, our discriminant probability is closer to 1. Hence, for the same discriminant distance, our approach will have a closer probability difference requirement to the source domain data. For example, assume that the image-level and instance-level discriminant probabilities of the source domain are $p_s = 0.1$ and $p_{s,j} = 0.11$, respectively, while they are $p_t = 0.8$ and $p_{t,j} = 0.88$ on the target domain. In this example, it is obvious that the discriminant distances in the source domain and the target domain are equal, but the discriminant probabilities in the source domain are closer according to the traditional Euclidean distance. Through this imbalanced optimization, we force the consistency of the discriminant in the labeled source domain to be accurate enough to extract useful features while relaxing the limit of discrimination in the unlabeled target domain to reduce the harm caused by incorrect instance extraction.

It is worth noting that our IPC loss only serves as a guide in the early stages of training. At the early stage of training, the discriminator can better determine whether the features come from the source or the target domain and thus perform effective alignment for different domains. However, at the later stage of training, after adversarial training, the feature extractor is better at extracting the public features, and the discriminator is harder to distinguish whether the extracted features come from the source or target domain, so the discriminant probability from source and target domains converges to 0.5. At that time, since the discriminant distances of features from the source and target domains are approximately the same, the IPC loss will treat the features from the source and target domains in a balanced manner. This is what we want because adversarial learning converges in the case where the network learns the common features, avoiding the damage to the network from the extraction of invalid features in the target domain in the early stages of training. After convergence, the IPC is more balanced to the source and target domains, giving the information in the target domain the correct migration effect.

Algorithm 1 Training DA Model With IDA Approach

```

for training iterations do
  Sample mini-batch data  $x_i^S, x_i^T$  from  $\mathcal{D}_S, \mathcal{D}_T$ 
  Extract image-level feature  $f_{\text{img}}(x_i^S)$  and  $f_{\text{img}}(x_i^T)$ 
  Extract instance-level feature  $f_{\text{ins}}(x_i^S)$  and  $f_{\text{ins}}(x_i^T)$ 
  for  $f_{\text{ins}}(x_i^S)$  do
    Calculate the classification loss  $L_{\text{cls}}$ 
    Calculate the bounding box regression loss  $L_{\text{reg}}$ 
  end for
  for paired  $f_{\text{img}}(x_i)$  and  $f_{\text{ins}}(x_i)$  do
    Calculate the domain discriminant probability  $p_i^{(u,v)}$ 
    and  $p_{i,j}$ 
    Calculate the image-level and instance-level loss  $L_{\text{img}}$ 
    and  $L_{\text{ins}}$ 
    for paired  $p_i^{(u,v)}$  and  $p_{i,j}$  do
      Calculate the IPC loss  $L_{\text{ipc}}$ 
    end for
  end for
  Update network parameters by backpropagation
end for

```

D. Total Objective

Our structure is improved on the DA model, and the detection network is Faster R-CNN. The loss of the detection part consists of two parts, which we denote as

$$L_{\text{det}} = L_{\text{cls}} + L_{\text{reg}} \quad (12)$$

where L_{cls} represents classification loss, and L_{reg} represents coordinate regression loss (detailed in [35]).

Overall, the final loss function in the training stage is the weighted sum of each loss function, which denotes as

$$L = L_{\text{det}} + \lambda(L_{\text{img}} + L_{\text{ins}}) + \alpha L_{\text{ipc}} \quad (13)$$

where λ and α are hyperparameters. In practice, we adjust the hyperparameters so that the order of magnitude of each loss is as consistent as possible to balance the influence of each part on the networks. The overall training data flow is detailed in algorithm 1.

The adversarial learning between discriminator and backbone networks is implemented by a GRL layer, and the standard SGD optimizer can optimize the loss of the whole network. During training, we input image pairs, with each mini-batch consisting of a labeled source domain image and an unlabeled target domain image. In the testing stage, the discrimination and domain classifier consistency modules are removed and only the Faster R-CNN detection network is used.

III. EXPERIMENTS

In this section, we first provide details about the dataset used and experimental setup, then outline the evaluation criteria for the experiments. We compare our method with several state-of-the-art unsupervised domain adaptation methods and indicate inspiring results on selected datasets. Additionally, we perform ablation experiments to confirm the efficacy of our proposed approach. Finally, we conduct parameter experiments and provide an analysis of the proposed algorithm.

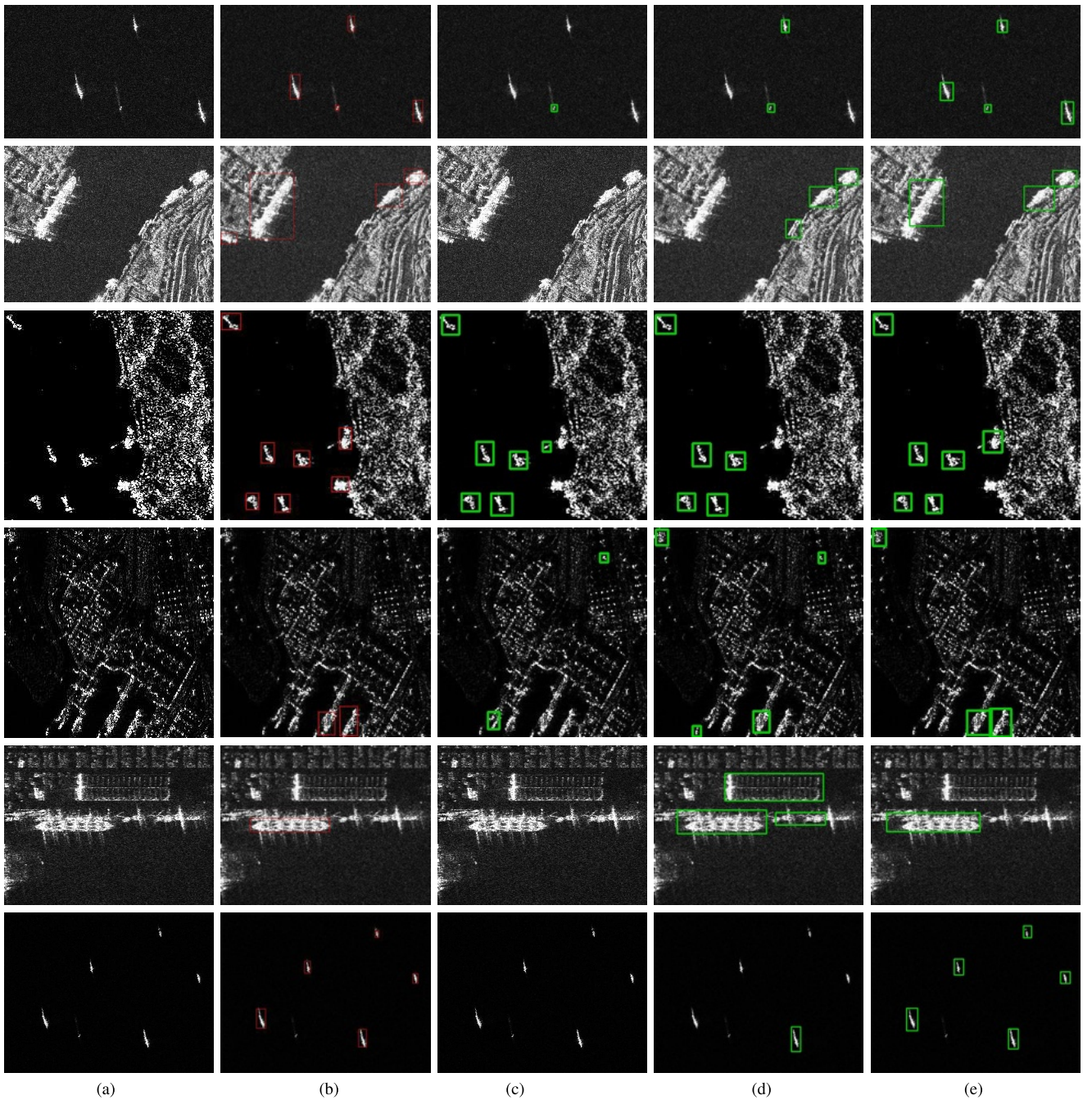


Fig. 3. Visualization of the detection results. The first two rows are the results of LEVIR to SSDD and the second two rows are LEVIR to SSD, the last two rows are HRSC to SSDD. (a) Input image, (b) ground truth, (c) detection results of the Faster R-CNN (source only), (d) detection results of the DA model, and (e) our detection results. Our method detects ships in some complex scenes and avoids the wrong judgment of some ship-like targets under SAR images.

A. Description of the Dataset and Experimental Settings

In this study, we utilize the visible images from “LEARNING, Vision and Remote sensing laboratory (LEVIR)” [37] and “High-Resolution Optical Satellite Images With Complex Backgrounds (HRSC)” [38] as the source domain and the SAR images from SAR Ship Detection Dataset (SSDD) [2] and SAR-Ship-Datasets [39] (SSD) as the target domains. These datasets all consist of images and their manual annotations. We make full use of the data from the source domain LEVIR while using only images in the training stage and

images and labels in the testing stage for the target domain data.

LEVIR contains 22 000 images sourced from Google Earth, with image sizes of 800×600 and resolutions ranging from 0.2 to 1.0 m per pixel. The dataset covers instances of 3 categories: airplanes (4724 instances), ships (3025 instances), and oil pots (3279 instances). In our experiments, we focus only on the 805 images containing ships and split the data into training, validation, and test sets in a 7:2:1 ratio.

HRSC contains 1061 images of six ship ports sourced from Google Earth, with image sizes of 300×300 to

1500 × 900. The dataset only covers 2976 ship instances, but these instances are subdivided into 19 different subcategories. In our experiments, we only focus on the broad category of ships and split the data into training, validation, and test sets in a 7:2:1 ratio.

The SSDD dataset contains 2456 ships in 1160 images from Radarsat-2, TerraSAR-X, and Sentinel-1 sensors. The polarization methods used include HH, VV, VH, and HV. The resolution of the images in the dataset ranges from 1 to 15 m per pixel and includes ships of different sizes and materials, both inshore and offshore. The training, validation, and testing sets are divided similar to the LEVIR dataset.

The SSD was created using 102 Chinese Gaofen-3 images and 108 Sentinel-1 images, with an image size of 256 × 256 and a resolution ranging from 3 to 25 m per pixel. It contains 43 819 images and 59 535 ship instances. In our experiments, we use 4745 images containing ships. The training and validation sets consist of 10% each of the images, while the rest are used for testing.

All experiments are implemented using PyTorch. Each mini-batch contains two images, one randomly selected from the source domain and another from the target domain. The pre-trained ResNet-101 from ImageNet is used as the backbone for the Faster R-CNN training. The parameters of the model are optimized using SGD with a weight decay of 0.0005 and a momentum of 0.9. The training process consists of 10 epochs. For the first 5 epochs, the learning rate is set to 0.002, then it is decayed to 0.0002 for the last 5 epochs. We select the model parameters that performs best on the validation set for testing. The total loss (13) is balanced by setting the hyperparameters to $\lambda = 0.1, \alpha = 1, \beta = 0.7$. During training and testing, regression boxes with instance intersection over union (IoU) greater than 0.5 are considered positive samples.

B. Evaluation Criteria

Following other SAR object detection tasks [40], [41], [42], we utilize average precision (AP) to evaluate the performance of our proposed model. For a given sample, it can be either a positive or negative sample, and our detection result can be either true or false. There are four possible cases: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Precision and recall can be defined based on these cases. Precision is a metric that measures the accuracy of positive predictions, expressed as the proportion of TP predictions to all positive samples. Recall, on the other hand, measures the proportion of actual positive samples that were correctly identified by the model

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (14)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (15)$$

Through our model, we can get a unique precision and recall value pair corresponding to a specific IoU threshold. By incrementally changing the IoU threshold from 0 to 1 and calculating precision and recall, we can plot a curve with precision on the vertical axis and recall on the horizontal axis.

The AP summarizes this precision–recall curve into a single value, representing the mean of all precisions. The definition of AP is as follows:

$$\text{AP} = \int_0^1 P(r) dr \quad (16)$$

$P(\cdot)$ is the function that represents the pair of precision and recall. The value of AP ranges from 0 to 1, where a higher value closer to 1 indicates better performance.

C. Experimental Results

To verify the effectiveness of our proposed model, we conduct experiments under four datasets to compare with several state-of-the-art domain adaptation algorithms: DA [32], SWDA [43], VDD-DAOD [44], CR [45], and HTC [46]. In all fairness, we assign the same data set partitioning to different algorithms. In Table I, we report the AP (%) on the test sets for various compared algorithms including our method. The second and third columns of Table I show the results when the source is LEVIR and the target domain is SSDD and SSD, respectively. And the fourth column shows the results that the source domain is HRSC and the target domain is SSDD. On SAR ship detection, our method achieves an improvement of 0.8%–8.11% over the state-of-the-art algorithms.

Comparing the performance of Oracle and Faster R-CNN, the mismatch between visible and SAR images causes the test results to be reduced to less than half of the similar distribution. The DA model using the domain adaptation method showed improvement in the AP, but there was still a noticeable gap with the upper bound. SWDA used strength-weakness alignment between instance- and image-level features to further enhance the detection performance, and our approach was also inspired by this imbalanced method. CR employed the technique of categorical regularization, which had an improvement of about 2% based on DA and SWDA. VDD-DOAD decoupled features from source and target domains, which is more appropriate for significant domain distance. HTC optimized the discriminator, which in turn affected the feature extractor. Our IDA approach also optimizes the discriminant network, leading to improved robustness of the backbone. Overall, the experimental results validate the superiority of our proposed model.

D. Further Analysis

1) *Ablation Study*: To explore the impact of the domain classifier consistency module and the significance of our proposed IDA approach, we design the following ablation experiments. Following the DA model, we utilize the l_2 loss for consistency, which can guarantee discriminant consistency in Euclidean space. Furthermore, we also employ the l_1 loss, which has a larger gradient than the l_2 loss. We also test the scenario where the domain classifier consistency module is removed, giving the discriminator more freedom in making choices. Table II shows the domain adaptation results of these implementations on LEVIR to SSDD and SSD, and HRSC to SSDD.

As Table II shows, we find that the results from using l_2 and l_1 loss for the domain classifier consistency module in

TABLE I

QUANTITATIVE RESULTS FOR SHIP DOMAIN DETECTION ADAPTING FROM LEVIR TO TWO DIFFERENT SAR DATASETS AND HRSC TO SSDD. FASTER R-CNN [35] TRAINED ONLY ON THE SOURCE DOMAIN, AND ORACLE TRAINED ON LABELED TARGET IMAGES USING FASTER R-CNN. WE REPORT THE AP RESULTS ON THE TARGET DOMAIN TEST SETS AND TRAINING TIME FOR VARIOUS COMPARISON METHODS. WE SELECT THE TRAINING PROCESS FROM HRSC TO SSDD AND REPORT THE AVERAGE TRAINING TIME PER 100 STEPS

Method	Datasets			Training Time (s/100 steps)
	LEVIR → SSDD	LEVIR → SSD	HRSC → SSDD	
Faster R-CNN	37.48%	29.74%	9.13%	20.25
DA	50.21%	47.10%	41.92%	29.81
SWDA	51.48%	47.32%	39.30%	35.57
VDD-DAOD	51.47%	51.98%	42.66%	57.40
CR-DA	53.74%	49.85%	43.26%	25.39
CR-SWDA	54.18%	49.66%	42.36%	29.47
HTCN	53.32%	50.26%	42.55%	43.85
Ours	55.10%	52.78%	47.41%	29.65
Oracle	88.81%	78.04%	88.81%	-

TABLE II

RESULTS OF VARIOUS CONSISTENCY LOSS IN ABLATION STUDIES

Method	Datasets		
	LEVIR → SSDD	LEVIR → SSD	HRSC → SSDD
DA (cst l_2 loss)	50.21%	47.10%	41.92%
DA (cst l_1 loss)	50.42%	45.06%	42.10%
DA w/o cst loss	50.31%	50.47%	39.97%
Ours IDA loss	55.10%	52.78%	47.41%

the LEVIR to SSDD domain adaptation have little difference. However, the AP resulting from l_1 loss shows a 2% decrease on LEVIR to SSD compared with l_2 loss. Indicating that the Euclidean distance is superior to the Manhattan distance for consistency. The scenario where the consistency loss is removed is also tested, and surprisingly, it does not have any impact on the AP of the LEVIR to SSDD domain adaptation, but the AP increases for the LEVIR to SSD domain adaptation. On the other hand, in the HRSC to SSDD experiment, the removal of the consistency loss term leads to a decrease in detection.

We analyze that the reason for this phenomenon is the difference in the optimization direction of the network led by the consistent loss under different distribution differences. In SAR and visible scenes with large distribution differences, the Euclidean distance treats the source and target domains in a balanced way, so the network learns features that are neither source nor target domains, and thus may not be as effective as removing the consistent term. In contrast, using our IPC loss, we unbalance the direction of optimization so that the network learns the common features better.

2) *Visualization Experiment*: To observe whether our model is aligned in the feature space, we randomly select 200 images from SSDD and SSD test sets and visualize the output feature of ResNet-101. For a clearer visualization, we reduce the high-dimensional features to two dimensions through the t -SNE [47]. As shown in Fig. 4, we compare the visualization results of the two groups of tasks on source only and our model. When the domain adaptation method is not used,

the features extracted from the source domain and the target domain exist in a gap, which leads to a rapid drop in the accuracy when the model trained on the source domain is tested on the target domain. On the other hand, using our model to mix the features of the source domain and the target domain well, allowing the subsequent detection head to accurately localize the target.

We visualize the final detection results, as shown in Fig. 3, we compare source only and DA model with our approach. Our method not only detects the existing target more effectively but also alleviate the misjudgment in virtual scenes.

3) *Influence of IoU Threshold*: In target detection tasks, the IoU between the prediction and ground truth is a criterion for measuring positive or negative samples. When the calculated IoU is greater than our specified threshold, the prediction is classified as a positive sample, otherwise it is a negative sample. Therefore, a higher threshold of IoU means stricter requirements for positive samples. Then we explore the influence of IoU thresholds on our detection results.

Fig. 5 illustrates the performance of various models (Source only, DA model, and ours) in different domain adaptation tasks as the IoU threshold varies. It is no longer meaningful since the AP obtained when the threshold value is greater than 0.75 is close to 10%. Therefore, we only plot the results with a threshold of 0.5–0.75. We find that as the threshold increases, the AP value decreases close to linearly. And our model has achieved the best results when the AP value greater than 20%, illustrating the effectiveness of our approach for accurate localization.

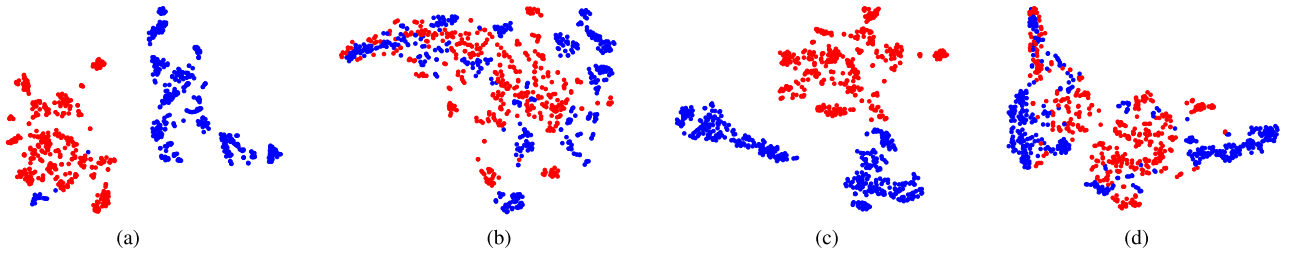


Fig. 4. Visualization of the extracted features by the backbone is presented using t-SNE [47] to compress the features into two dimensions. The red points indicate features from the source domain, while the blue points are from the target domain. (a) and (b) Features of Faster R-CNN and our method on the LEVIR to SSDD datasets, while (c) and (d) features of Faster R-CNN and our method on the LEVIR to SSD datasets. The comparison illustrates that our method reduces the distribution of features, which helps the backbone extract common features to both domains.

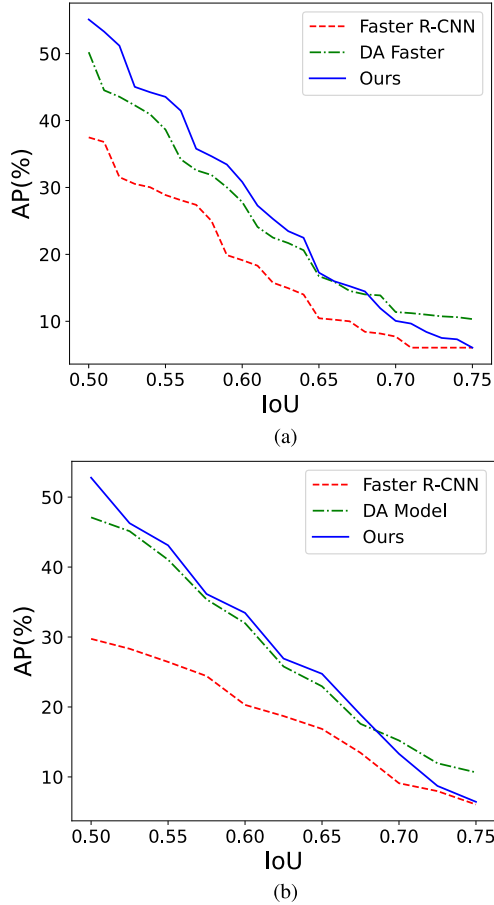


Fig. 5. Detection accuracy with the various IoU thresholds on different adaptation task. In both sets of domain adaptation tasks, detection accuracy decreases similarly linearly with increasing threshold. (a) LEVIR→SSDD. (b) LEVIR→SSD.

4) *Training Time Statistics:* Algorithmic performance improvements may come at the cost of computational time. However, the IDA method in this article does not add redundant structure to the baseline work, and the IPC loss only changes the gradient backpropagation process. Table I shows the training time of our method compared to other methods on HRSC to SSDD. There is no significant increase in our average training time compared to the DA model. In addition to the CR model streamlining the structure of DA, our method still dominates in training time.

5) *Analysis of Failed Detection:* Fig. 6 shows several failures encountered by our model. These failures can be classified into two main categories: 1) the inability to detect

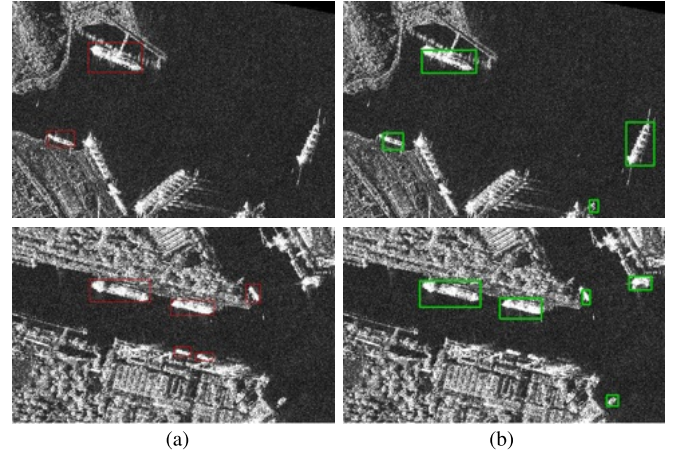


Fig. 6. Demonstration of the failure on LEVIR→SSDD, in the case of complex scenarios, there may still be missed detections and wrong detections. (a) GT. (b) Ours.

certain existing targets, especially in complex scenes like ships docked on the shore; and 2) the erroneous identification of the ship-like object as the actual ship in complex scenes. Additionally, we note that other domain adaptation models face similar challenges in detecting targets when transitioning from visible to SAR images. While our model does mitigate these errors to some extent, addressing these challenges will be the focus of our future research.

IV. CONCLUSION

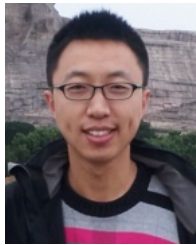
SAR images are used for detecting suspicious ships due to their immunity to weather interference. However, the imaging principle of SAR images makes it challenging to obtain a large amount of labeled data. To address this issue, the domain adaptation method can migrate information from the source domain to the target domain, so the unlabeled SAR image can borrow information from the labeled visible image. In this work, we propose an IDA approach for domain classifier consistency embedded in a ship detection model in SAR images based on domain adaptation. Our essential contribution is to deal with the negative migration problem when the distribution of two domains is not similar, particularly when the source domain is the visible image and the target domain is the SAR image. The IDA approach performs strict alignment on the source domain while applying soft alignment on the target domain by utilizing IPC loss, which improves the robustness of feature extraction networks. Meanwhile, we theoretically prove the effectiveness of our proposed IDA approach. And

we validate the superiority of our proposed model on two sets of domain adaptation tasks on three sets of datasets. Our approach achieves 0.8%–8.11% improvement in AP metrics on the state-of-the-art domain adaptation works. In addition, we discuss the impact of different distance measures on the alignment.

REFERENCES

- [1] J. Liu, P. Ren, X. Lyu, and C. Grecos, "A Gamma-Log net for oil spill detection in inhomogeneous SAR images," *Remote Sens.*, vol. 14, no. 16, p. 4074, Aug. 2022, doi: [10.3390/rs14164074](https://doi.org/10.3390/rs14164074).
- [2] Z. Cui, Q. Li, Z. Cao, and N. Liu, "Dense attention pyramid networks for multi-scale ship detection in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8983–8997, Nov. 2019.
- [3] D. Wang et al., "BO-DRNet: An improved deep learning model for oil spill detection by polarimetric features from SAR images," *Remote Sens.*, vol. 14, no. 2, p. 264, Jan. 2022, doi: [10.3390/rs14020264](https://doi.org/10.3390/rs14020264).
- [4] Y. Liu, Q. Li, Y. Yuan, Q. Du, and Q. Wang, "ABNet: Adaptive balanced network for multiscale object detection in remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5614914, doi: [10.1109/TGRS.2021.3133956](https://doi.org/10.1109/TGRS.2021.3133956).
- [5] L. Ouyang, L. Fang, and X. Ji, "Multigranularity self-attention network for fine-grained ship detection in remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 9722–9732, 2022.
- [6] H. Lin, Z. Shi, and Z. Zou, "Fully convolutional network with task partitioning for inshore ship detection in optical remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1665–1669, Oct. 2017.
- [7] Q. Huang, W. Zhu, Y. Li, B. Zhu, T. Gao, and P. Wang, "Survey of target detection algorithms in SAR images," in *Proc. IEEE 5th Adv. Inf. Technol., Electron. Autom. Control Conf. (IAEAC)*, vol. 5, 2021, pp. 1756–1765.
- [8] M. C. Burl, G. J. Owirka, and L. M. Novak, "Texture discrimination in synthetic aperture radar imagery," in *Proc. 23rd Asilomar Conf. Signals, Syst. Comput.*, Oct./Nov. 1989, pp. 399–400.
- [9] Y. Huang and F. Liu, "Detecting cars in VHR SAR images via semantic CFAR algorithm," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 6, pp. 801–805, Jun. 2016.
- [10] W. Yu, Y. Wang, H. Liu, and J. He, "Superpixel-based CFAR target detection for high-resolution SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 5, pp. 730–734, May 2016.
- [11] B. Hou, X. Chen, and L. Jiao, "Multilayer CFAR detection of ship targets in very high resolution SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 4, pp. 811–815, Apr. 2015.
- [12] Y. Wang, Q. Yu, W. Lv, and W. Yu, "Coastline detection in SAR images using multi-feature and SVM," in *Proc. 4th Int. Congr. Image Signal Process.*, vol. 3, Oct. 2011, pp. 1227–1230.
- [13] Z. Cao, Z. Cui, Y. Fan, and Q. Zhang, "SAR automatic target recognition using a hierarchical multi-feature fusion strategy," in *Proc. IEEE Globecom Workshops*, Dec. 2012, pp. 1450–1454.
- [14] T. Li, Z. Liu, R. Xie, and L. Ran, "An improved superpixel-level CFAR detection method for ship targets in high-resolution SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 1, pp. 184–194, Jan. 2018.
- [15] X. Yang, F. Bi, Y. Yu, and L. Chen, "An effective false-alarm removal method based on OC-SVM for SAR ship detection," in *Proc. IET Int. Radar Conf.*, Oct. 2015, pp. 1–4, doi: [10.1049/cp.2015.1368](https://doi.org/10.1049/cp.2015.1368).
- [16] G. Qu, Q. Yu, and Y. Wang, "An improved method for SAR image coastline detection based on despeckling and SVM," in *Proc. IET Int. Radar Conf.*, Apr. 2013, pp. 1–6, doi: [10.1049/cp.2013.0268](https://doi.org/10.1049/cp.2013.0268).
- [17] K. Ji, X. Leng, H. Wang, S. Zhou, and H. Zou, "Ship detection using weighted SVM and M-CHI decomposition in compact polarimetric SAR imagery," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2017, pp. 890–893.
- [18] G. Wei, Q. Qingwen, J. Lili, and Z. Ping, "A new method of SAR image target recognition based on AdaBoost algorithm," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2008, pp. III-1194–III-1197.
- [19] R. Min, H. Quan, Z. Cui, Z. Cao, Y. Pi, and Z. Xu, "SAR target detection using AdaBoost via GPU acceleration," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2019, pp. 1180–1183.
- [20] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
- [21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Dec. 2016, pp. 770–778.
- [22] Q. Li, R. Min, Z. Cui, Y. Pi, and Z. Xu, "Multiscale ship detection based on dense attention pyramid network in SAR images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2019, pp. 5–8.
- [23] W. Shi, Z. Hu, H. Liu, S. Cen, J. Huang, and X. Chen, "Ship detection in SAR images based on adjacent context guide fusion module and dense weighted skip connection," *IEEE Access*, vol. 10, pp. 134263–134276, 2022.
- [24] Y. Zhao, L. Zhao, Z. Liu, D. Hu, G. Kuang, and L. Liu, "Attentional feature refinement and alignment network for aircraft detection in SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5220616, doi: [10.1109/TGRS.2021.3139994](https://doi.org/10.1109/TGRS.2021.3139994).
- [25] Y. Zhu, X. Sun, W. Diao, H. Li, and K. Fu, "RFA-Net: Reconstructed feature alignment network for domain adaptation object detection in remote sensing imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 5689–5703, 2022.
- [26] J. Zhang, S. Li, Y. Dong, B. Pan, and Z. Shi, "Hierarchical similarity alignment for domain adaptive ship detection in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5240611, doi: [10.1109/TGRS.2022.3227626](https://doi.org/10.1109/TGRS.2022.3227626).
- [27] C. Zhu, D. Zhao, Z. Liu, and Y. Mao, "Hierarchical attention for ship detection in SAR images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Sep. 2020, pp. 2145–2148.
- [28] G. Cheng et al., "Prototype-CNN for few-shot object detection in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5604610.
- [29] J. Peng, W. Sun, L. Ma, and Q. Du, "Discriminative transfer joint matching for domain adaptation in hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 6, pp. 972–976, Jun. 2019.
- [30] M. Wang and W. Deng, "Deep visual domain adaptation: A survey," *Neurocomputing*, vol. 312, pp. 135–153, Oct. 2018.
- [31] Y. Shi, L. Du, and Y. Guo, "Unsupervised domain adaptation for SAR target detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 6372–6385, 2021.
- [32] Y. Chen, W. Li, C. Sakaridis, D. Dai, and L. Van Gool, "Domain adaptive faster R-CNN for object detection in the wild," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3339–3348.
- [33] Y. Song, J. Li, P. Gao, L. Li, T. Tian, and J. Tian, "Two-stage cross-modality transfer learning method for military-civilian SAR ship recognition," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, doi: [10.1109/LGRS.2022.3162707](https://doi.org/10.1109/LGRS.2022.3162707).
- [34] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2242–2251.
- [35] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.
- [36] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by back-propagation," in *Proc. 32nd Int. Conf. Mach. Learn. (ICML)*, vol. 37, Jul. 2015, pp. 1180–1189.
- [37] Z. Zou and Z. Shi, "Random access memories: A new paradigm for target detection in high resolution aerial remote sensing images," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1100–1111, Mar. 2018.
- [38] Z. Liu, H. Wang, L. Weng, and Y. Yang, "Ship rotated bounding box space for ship extraction from high-resolution optical satellite images with complex backgrounds," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 8, pp. 1074–1078, Aug. 2016.
- [39] Y. Wang, C. Wang, H. Zhang, Y. Dong, and S. Wei, "A SAR dataset of ship detection for deep learning under complex backgrounds," *Remote Sens.*, vol. 11, no. 7, p. 765, Mar. 2019, doi: [10.3390/rs11070765](https://doi.org/10.3390/rs11070765).
- [40] Z. Lin, K. Ji, X. Leng, and G. Kuang, "Squeeze and excitation rank faster R-CNN for ship detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 5, pp. 751–755, May 2019.
- [41] S. Wei, X. Zeng, Q. Qu, M. Wang, H. Su, and J. Shi, "HRSID: A high-resolution SAR images dataset for ship detection and instance segmentation," *IEEE Access*, vol. 8, pp. 120234–120254, 2020.
- [42] Q. An, Z. Pan, L. Liu, and H. You, "DRBox-v2: An improved detector with rotatable boxes for target detection in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8333–8349, Nov. 2019.
- [43] K. Saito, Y. Ushiku, T. Harada, and K. Saenko, "Strong-weak distribution alignment for adaptive object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 6949–6958.

- [44] A. Wu, R. Liu, Y. Han, L. Zhu, and Y. Yang, "Vector-decomposed disentanglement for domain-invariant object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9322–9331.
- [45] C.-D. Xu, X.-R. Zhao, X. Jin, and X.-S. Wei, "Exploring categorical regularization for domain adaptive object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11721–11730.
- [46] C. Chen, Z. Zheng, X. Ding, Y. Huang, and Q. Dou, "Harmonizing transferability and discriminability for adapting object detectors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 8866–8875.
- [47] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 2579–2605, Nov. 2008.



Bin Pan (Member, IEEE) received the B.S. and Ph.D. degrees from the School of Astronautics, Beihang University, Beijing, China, in 2013 and 2019, respectively.

Since 2019, he has been an Associate Professor with the School of Statistics and Data Science, Nankai University, Tianjin, China. His research interests include machine learning, remote sensing image processing, and multiobjective optimization.



Zhehao Xu received the B.S. degree from the College of Mathematics, Harbin Institute of Technology, Harbin, China, in 2021, and the M.S. degree from the School of Statistics and Data Science, Nankai University, Tianjin, China, in 2023.

His research interests include machine learning, object detection, and image inpainting.



Tianyang Shi received the B.S. and M.S. degrees from the School of Astronautics, Beihang University, Beijing, China, in 2016 and 2019, respectively.

He is currently a Researcher with ByteDance, Beijing. He has published more than 20 peer-reviewed papers, including top-tier conferences, such as the IEEE Conference on Computer Vision and Pattern Recognition and the IEEE International Conference on Computer Vision. His research interests include image processing, deep learning, and their application in games. His homepage is <https://www.shitianyang.tech/>.



Tao Li received the Ph.D. degree in computer science from Nankai University, Tianjin, China, in 2007.

He is currently with the College of Computer Science, Nankai University, as a Professor. His research interests include heterogeneous computing, machine learning, and the Internet of Things.

Dr. Li is a Distinguished Member of the China Computer Federation (CCF).



Zhenwei Shi (Member, IEEE) received the Ph.D. degree in mathematics from the Dalian University of Technology, Dalian, China, in 2005.

He was a Post-Doctoral Researcher with the Department of Automation, Tsinghua University, Beijing, China, from 2005 to 2007. He was a Visiting Scholar with the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL, USA, from 2013 to 2014. He is currently a Professor and the Dean of the Image Processing Center, School of Astronautics, Beihang University, Beijing.

He has authored or coauthored over 200 scientific articles in refereed journals and proceedings, including IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), and the IEEE International Conference on Computer Vision (ICCV). His research interests include remote sensing image processing and analysis, computer vision, pattern recognition, and machine learning.

Dr. Shi serves as an Editor for IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the *Pattern Recognition*, the *ISPRS Journal of Photogrammetry and Remote Sensing*, and the *Infrared Physics and Technology*. His personal website is <http://levir.buaa.edu.cn/>.