*remote sensing*

MDPI

*Article*

# SVIADF: Small Vessel Identification and Anomaly Detection Based on Wide-Area Remote Sensing Imagery and AIS Data Fusion

**Lihang Chen, Zhuhua Hu *, Junfei Chen and Yifeng Sun**

School of Information and Communication Engineering, Hainan University, Haikou 570228, China; 23210810000036@hainanu.edu.cn (L.C.); 22220854000129@hainanu.edu.cn (J.C.); 22210810000036@hainanu.edu.cn (Y.S.)
* Correspondence: eagler_hu@hainanu.edu.cn

**Abstract:** Small target ship detection and anomaly analysis play a pivotal role in ocean remote sensing technologies, offering critical capabilities for maritime surveillance, enhancing maritime safety, and improving traffic management. However, existing methodologies in the field of detection are predominantly based on deep learning models with complex network architectures, which may fail to accurately detect smaller targets. In the classification domain, most studies focus on synthetic aperture radar (SAR) images combined with Automatic Identification System (AIS) data, but these approaches have significant limitations: first, they often overlook further analysis of anomalies arising from mismatched data; second, there is a lack of research on small target ship classification using wide-area optical remote sensing imagery. In this paper, we develop SVIADF, a multi-source information fusion framework for small vessel identification and anomaly detection. The framework consists of two main steps: detection and classification. To address challenges in the detection domain, we introduce the YOLOv8x-CA-CFAR framework. In this approach, YOLOv8x is first utilized to detect suspicious objects and generate image patches, which are then subjected to secondary analysis using CA-CFAR. Experimental results demonstrate that this method achieves improvements in Recall and F1-score by 2.9% and 1.13%, respectively, compared to using YOLOv8x alone. By integrating structural and pixel-based approaches, this method effectively mitigates the limitations of traditional deep learning techniques in small target detection, providing more practical and reliable support for real-time maritime monitoring and situational assessment. In the classification domain, this study addresses two critical challenges. First, it investigates and resolves anomalies arising from mismatched data. Second, it introduces an unsupervised domain adaptation model, Multi-CDT, for heterogeneous multi-source data. This model effectively transfers knowledge from SAR–AIS data to optical remote sensing imagery, thereby enabling the development of a small target ship classification model tailored for optical imagery. Experimental results reveal that, compared to the CDTrans method, Multi-CDT not only retains a broader range of classification categories but also improves target domain accuracy by 0.32%. The model extracts more discriminative and robust features, making it well suited for complex and dynamic real-world scenarios. This study offers a novel perspective for future research on domain adaptation and its application in maritime scenarios.

**Keywords:** remote sensing imagery; Automatic Identification System (AIS); vessel identification; multi-modal information fusion; anomaly assessment



Academic Editor: Andrzej Stateczny

Received: 23 January 2025
Revised: 21 February 2025
Accepted: 21 February 2025
Published: 28 February 2025

**Citation:** Chen, L.; Hu, Z.; Chen, J.; Sun, Y. SVIADF: Small Vessel Identification and Anomaly Detection Based on Wide-Area Remote Sensing Imagery and AIS Data Fusion. *Remote Sens.* **2025**, *17*, 868. https://doi.org/10.3390/rs17050868

## 1. Introduction

With the accelerating pace of global economic integration, the significance of the oceans in international trade, energy development, and fisheries resource utilization has increasingly come to the forefront. The ocean not only serves as a vital passage for international shipping and trade but also as a key source of offshore oil, natural gas, and fisheries resources. In China, as the construction of the Hainan Free Trade Port progresses steadily, maritime traffic and economic activities in the surrounding waters have intensified, further highlighting the strategic value of the ocean in the nation's economic and environmental governance. However, the frequent and diversified nature of maritime economic activities has also brought to the forefront issues related to maritime security and environmental protection. As key carriers of maritime transport and economic activity, vessels play a pivotal role in driving international trade and resource development while simultaneously serving as primary targets for monitoring illegal maritime activities, ensuring regional security, and protecting the marine environment.

Against this backdrop, ensuring the effective identification and anomaly monitoring of small vessels has become increasingly important. Existing research primarily focuses on vessel anomaly detection based on AIS (Automatic Identification System) trajectory data [1,2]. However, monitoring methods that rely on AIS signals faces significant challenges. Malicious actors may deactivate or tamper with AIS data to evade surveillance and monitoring [3,4]. To counter this, some studies have incorporated shore-based radar trajectories to supplement AIS tracking [5–8]. Yet, shore-based radar is limited to nearshore areas and struggles to cover vast and complex high seas area. To address the blind spots inherent in AIS monitoring, remote sensing technologies have emerged as a powerful supplementary tool. Remote sensing imagery offers all-weather, all-encompassing monitoring of the sea surface, particularly in adverse weather conditions or when AIS data are unavailable. In such instances, remote sensing images can still provide effective vessel information. Consequently, some existing studies have sought to integrate remote sensing imagery and AIS data to enable vessel monitoring across a broader range of maritime areas [9–14], thus offering a more comprehensive information base for detecting illicit activities and supporting maritime safety management. However, the majority of vessel recognition methods in current research rely on successful matches between AIS data and remote sensing imagery, often neglecting the need for further analysis of anomalies arising from mismatched or failed matches. This gap presents a significant challenge for the comprehensive and reliable monitoring of maritime activities.

In this study, high-resolution wide-area optical remote sensing images and AIS data acquired from China's Hainan Satellite 1 are fused to detect and classify small target ships and make further judgment on abnormal conditions to realize the identification of small target ships and their abnormal judgment so as to strengthen the monitoring of abnormal events and support joint prevention and control measures, as well as explore the challenges and solutions in this field.

In terms of vessel detection, the field can be divided into SAR remote sensing imagery and optical remote sensing imagery. In the SAR domain, traditional CFAR algorithms and their derivatives [15–18] have been found to perform well in small target detection but are limited in variable environments. Subsequent improvements have been proposed [19–24]. In the field of optical remote sensing images, relevant datasets [25–28] provide the basis for deep learning. However, most of the remote sensing images of these datasets are concentrated in the coastal position, and the resolution is small, which does not meet the monitoring requirements of wide sea areas in our practical application. Therefore, we constructed a wide-area optical remote sensing dataset using imagery captured by the Hainan Satellite 1 [29]. In addition, in the detection model, the current target detection

methods based on deep learning mostly use deep network structure. Although the multi-level feature fusion method is used to improve the detection performance of small target objects [30–37], insufficient attention is still paid to shallow features, which easily leads to unsatisfactory detection results of small target ships. In the field of ship classification, many studies have focused on the combination of SAR images and AIS data and have made significant progress [38–40]. However, there are still shortcomings in the field of small target ship classification of optical remote sensing images. The fundamental problem is that it is difficult to obtain sample labels for small target ship classification under wide-area optical remote sensing images. This problem can be solved by the unsupervised domain adaptation method [41–53]. Thus, the problem of insufficient data labels is alleviated. However, the existing research mainly focuses on the migration of multi-source homogeneous data or single-source heterogeneous data and rarely considers the fusion and migration of multi-source heterogeneous data. Traditional multi-source information fusion methods rely on manual feature extraction [54–56], whereas Transformer-based methods enhance flexibility and adaptability through self-attention mechanisms and end-to-end learning [57,58]. The current issues are as follows:

(1) The existing research lacks the identification of small target vessels based on the fusion of optical remote sensing images and AIS data, and there is a shortage of further discussion and judgment on the abnormal cases where the matching between remote sensing images and AIS information fails.
(2) In the task of small target vessel detection in remote sensing imagery, the appearance features of vessels are often confined to a very small spatial range (sometimes as low as 10 pixels), and typical deep learning-based object detection methods with deeper network structures may miss shallow-level targets.
(3) In the classification task of small target vessels in remote sensing imagery, current research primarily focuses on the classification of small target vessels in SAR imagery, with relatively little research in the field of optical remote sensing imagery. To address this, unsupervised domain adaptation transfer learning has been proposed, but it faces the challenge of multi-source data processing. Existing transfer learning research mainly focuses on the transfer of multi-source homogeneous data or single-source heterogeneous data, lacking research on the transfer of multi-source heterogeneous data.

In response to the key scientific and engineering issues raised, the main contributions of this paper can be summarized as follows:

(1) This study constructs a multi-source information fusion framework for small vessel identification and anomaly detection (SVIADF). It achieves the identification of small target vessels using optical remote sensing images and AIS information while also addressing further discussion and judgment on the abnormal cases where the matching fails.
(2) To address the issue of vessels having fewer pixels and less-distinct features in high-resolution, wide-area optical remote sensing imagery, the YOLOv8x-CA-CFAR is proposed, enabling the model to combine structure-based and pixel-based methods and significantly improve the model's Recall rate.
(3) To address the challenge of obtaining classified sample data from high-resolution, wide-area optical remote sensing imagery, unsupervised domain adaptation is proposed. Furthermore, the Multi-CDT classification model for multi-source heterogeneous data is introduced, transferring SAR-AIS data to optical remote sensing imagery to enhance model performance.

The remainder of this paper is organized as follows: Section 2 provides a further introduction to previous research in various fields. In Section 3, the proposed SVIADF is introduced first, and then the YOLOv8x-CA-CFAR and Multi-CDT algorithm models are introduced. Section 4 analyzes and discusses the performance of the proposed algorithm models through experimental results and offers future prospects. Finally, Section 5 concludes the paper.

## 2. Related Work

### *2.1. Small Target Vessels Anomaly Detection Task*

In Table 1, we present the relationships of existing anomaly detection studies.

**Table 1.** Relationship table of vessel anomaly detection algorithms.

| Research Area | Data Source | | | | Core Technology | Existing Research |
|---|---|---|---|---|---|---|
| | Radar Remote Sensing Images | Optical Remote Sensing Images | AIS | Other | | |
| Anomaly Detection | × | × | ✓ | × | Anomaly Trajectory Detection | CBVAD [1] MFGTN [2] |
| | × | × | ✓ | × | | STADS [5] RABD [6] RFSPA-CRCS [7] JCAPA-URCN [8] |
| | ✓ | × | ✓ | × | Data Matching and Fusion | SIFMON [9] FISAR [10] |
| | ✓ | ✓ | ✓ | × | | MSDF-MSA [11] |
| | × | ✓ | ✓ | × | | SAFIMON [12] |
| | ✓ | × | ✓ | × | | MARSPROM [13] CLASS-SAR [14] |

Most existing research focuses on the detection of vessel anomalies based on AIS (Automatic Identification System) trajectory data, utilizing vessels position information, speed, course, and other data to analyze whether vessel behavior conforms to normal navigation patterns [1] or to determine the type of vessel operations and then, in conjunction with regulatory knowledge, assess whether the behavior is anomalous [2].

However, AIS data may be subject to shutdown or tampering by humans [3,4], leading to blind spots and uncertainty in anomaly detection based on AIS data. To improve the comprehensiveness and accuracy of monitoring, some studies have attempted to supplement AIS trajectory detection with coastal radar trajectory data [5–8]. Coastal radar can locate vessels monitoring in nearshore areas by detecting the physical location trajectories of vessels and comparing them with AIS data, thereby identifying and correcting anomalies in AIS data.

Nevertheless, the application of coastal radar is limited by its coverage area, effectively monitoring vessels only in nearshore waters and struggling to effectively monitor vessels far from the coast. In order to make up for these limitations, in recent years, some studies have proposed to combine remote sensing images and AIS data for ship monitoring [9–14]. The ship target is identified from satellite images and matched with AIS data to identify the ship identity. The trajectory interpolation method based on AIS is designed to solve the problem of time alignment between remote sensing images and AIS information [9–12] and improve the reliability of matching.

However, the analysis of the abnormal situation after the failure of matching is insufficient in the current research. The failure of matching may not only be caused by non-human factors such as technical problems and environmental interference but also may involve abnormal behaviors such as illegal fishing and smuggling. Although it is

mentioned in [9] to report the abnormal information that fails to match, there is still a lack of in-depth judgment on the matching failure caused by artificially closing or tampering AIS information. The difficulty is that it is difficult for the high-seas area to be supplemented by other data sources, so it is particularly difficult to make further abnormal judgment in the case of matching failure. In [13], in addition to matching remote sensing images and AIS information, an attempt was made to predict the speed of the ship based on the wake in the image, while in [14], a classifier was trained based on AIS data, hoping to supplement the category prediction information. In general, in practical applications, we hope that in addition to matching remote sensing images and AIS information, more useful information can be predicted from limited data sources to better meet the needs of subsequent monitoring.

### *2.2. Small Target Vessels Detection Task*

In Table 2, we present the relationships of existing small target vessels detection studies.

**Table 2.** Relationship table of small target vessel detection algorithms.

| Research Area | Data Source | | | | Core Technology | Existing Research |
|---|---|---|---|---|---|---|
| | Radar Remote Sensing Images | Optical Remote Sensing Images | AIS | Other | | |
| Detection | ✓ | × | × | × | Constant False Alarm Rate | CFAR [15]<br>CA-CFAR [18]<br>Proposal-Based CFAR [20]<br>Faster R-CNN CFAR [21]<br>Haar-like CA-CFAR [22]<br>IP-CFAR [23]<br>SSD CFAR [24] |
| | × | ✓ | × | × | Multi-Scale Feature Fusion | FPN [30]<br>PAN [31]<br>Swin Transformer v2 [32]<br>PVT v2 [33]<br>Yolov8 [34]<br>SAFF-SDD [35]<br>CSDP-YOLO [36]<br>OptiShipNet [37] |
| | ✓ | × | ✓ | ✓ | Multi-Scale Feature Fusion | MuIT [54]<br>UniT [55]<br>ACDF-YOLO [56] |
| | ✓ | ✓ | ✓ | × | | |
| | × | ✓ | × | ✓ | | |

In the field of SAR imagery, the most widely applied traditional algorithm is the CFAR (Constant False Alarm Rate) and its derivatives [15–18]. CFAR methods are favored because they can dynamically adjust the detection threshold based on the statistical characteristics of background noise, adapting to target detection in complex backgrounds. CFAR algorithms are relatively simple and computationally efficient, making them suitable for large-scale applications with high real-time requirements. They demonstrate advantages in dealing with speckle noise and ground interference in SAR images.

However, traditional CFAR methods rely on statistical assumptions and a single-threshold strategy, which can be limited in low-signal-to-noise ratio environments [20,22]. Therefore, recent studies have combined CFAR with deep learning [21,23,24] to enhance its applicability and detection performance. The advantage of this combination is that deep learning effectively improves the target recognition ability in complex backgrounds by automatically extracting high-dimensional features, while CFAR enhances the robustness in low-SNR environments by adaptively adjusting the threshold. Compared to using CFAR alone, the combination of deep learning and CFAR can make up for the limitations of CFAR method in low-signal-to-noise ratio and clutter conditions, as well as improve the accuracy and reliability of detection.

In the domain of optical imagery, the detection accuracy of small target objects has been significantly improved thanks to the multi-scale feature fusion technologies of the FPN (Feature Pyramid Network) [30] and PAN (Path Aggregation Network) [31]. The latest deep learning models that combine multi-level features have shown excellent performance in small target vessels detection [32–37], and the fusion of multi-source information further enhances recognition accuracy and robustness [54–56].

However, deeper network structures may lead to the loss of small target features, necessitating a focus on the effective utilization of shallow features. Therefore, combining shallow and deep features to improve the accuracy and robustness of small object detection is an important direction of current research.

### *2.3. Small Target Vessels Classification Task*

In Table 3, we present the relationships of existing small target vessel classification studies.

**Table 3.** Relationship table of small target Vessel classification algorithms.

| Research Area | Data Source | | | | Core Technology | Existing Research |
|---|---|---|---|---|---|---|
| | Radar Remote Sensing Images | Optical Remote Sensing Images | AIS | Other | | |
| Classification | ✓ | × | ✓ | × | Data Matching and Automatic Annotation | LAVMON-DPS-AIS [38]<br>FUSAR-Ship [39]<br>MCEL [40] |
| | × | ✓ | ✓ | × | Not Yet Available | Not Yet Available |
| | ✓ | ✓ | × | ✓ | Unsupervised Domain Adaptation | DAN [41]<br>CORAL [42]<br>JAN [43]<br>TTL-ML [44]<br>Improved OSDA-BP [45]<br>GAN [46]<br>DANN [47,48]<br>AFN [49]<br>MAN [50]<br>MCD [51]<br>BSP [52]<br>CDTrans [53] |
| | × | × | × | ✓ | Multi-Source Information Fusion | MuIT [57]<br>UniT [58] |
| | ✓ | ✓ | ✓ | × | Multiple Sources | MSARCDSC [59] |

In the field of SAR imagery, the collaborative use of SAR images and AIS data is the primary approach, utilizing AIS data to annotate vessels in SAR images and providing supervised information. Studies have set a maximum tolerance distance of 2 km between SAR and AIS data [38], improving the reliability of matching and establishing a high-resolution SAR–AIS matched dataset called FUSAR-Ship [39], which provides an experimental basis and guidance for vessels classification research. Through multi-classifier ensemble learning and AIS data transfer methods [40], more complex vessels classification tasks have been achieved in SAR images.

In the domain of optical imagery, the research is relatively scarce due to the difficulty of obtaining labeled data for high-resolution, wide-area remote sensing images.

To address this issue, the proposal is to transfer SAR–AIS data to optical imagery, involving unsupervised domain adaptation and multi-source information fusion techniques.

Early unsupervised domain adaptation methods primarily reduced the distribution difference between source and target domains through metrics like Maximum Mean Discrepancy (MMD) [41–43], but their effectiveness was limited by sample imbalance and insufficient feature alignment, with subsequent improvements proposed [44,45]. Adversarial domain adaptation methods [46–50] can more precisely match feature distributions but suffer from training instability. Methods like the MCD [51] and BSP [52] offer different perspectives, being particularly suitable for feature boundary adjustment in specific tasks.

CDTrans [53] enhances domain alignment effects using the cross-attention mechanism of Transformers, improving model robustness.

In terms of multi-source information fusion, Transformer-based methods, through self-attention mechanisms and end-to-end learning [57,58], can flexibly handle multi-source information, capture global correlations, and do not require manual feature engineering, thus possessing stronger expressive power and adaptability.

In the task of unsupervised domain adaptation transfer learning for multi-source domains, while multi-source homogeneous problems can be addressed using distribution similarity indices (DSIs metrics) [59], the research on multi-source heterogeneous situations remains insufficient.

## 3. Proposed Method

Since multiple datasets or data sources are mentioned in this study, in order to facilitate understanding of the relationships and differences between these data sets or data sources, a detailed overview of them is given in Table 4, including data types, resolution, and other information.

In addition, in order to better show the research context of this paper, we briefly explain the application of these datasets or data sources in this research.

Firstly, HRSC2016 [25] and DOTA [26], which are optical remote sensing datasets, were not used in this study, aiming to comparatively illustrate that most of the existing optical remote sensing datasets are small-resolution nearshore remote sensing datasets, which are different from the situation in this study. HN01-SHIP [29] is a small target ship dataset constructed based on the remote sensing image of Hainan Satellite 1 in this study, which was used for small target ship detection and classification. AIR-SARShip-2.0 [60] was used to compare the differences between optical remote sensing images and SAR images, because this dataset has a ship labeling box for SAR images, which is convenient for subsequent slice comparison. FUSAR [39] was used as the source domain for multi-source heterogeneous unsupervised domain adaptation and transferred to our dataset built on Hainan Satellite 1. In addition, the method proposed in this study is not only applicable to Hainan Satellite 1 but also has a wide range of data source adaptation capabilities. Therefore, Sentinel-1 data were used to supplement the generality and adaptability of the verification framework for further demonstration.

**Table 4.** Datasets used and their key characteristics.

| Dataset Name | Data Source | Data Types | Image Resolution | Image Size Range | Reference Date | Feature Description |
|---|---|---|---|---|---|---|
| HRSC2016 [25] | Google Earth | Optical | 0.4∼2 m | 300∼1500 | 2016 | Focusing on ship detection; the images are diverse and suitable for maritime target detection. |
| DOTA [26] | Google Earth, CRESDA | Optical | 0.3∼2 m | 800∼4000 | 2018 | Used for target detection in aerial imagery, multi-scene, multi-scale, complex target. |
| HN01-SHIP [29] | Hainan Satellite 1 | Optical AIS | 5 m | 1024 | 2022 2023 | The resolution is large, the ship scale is small, and the background is complex. |
| AIR-SARShip-2.0 [60] | Gaofen-3 | SAR | 1 m and 3 m | 1000 | 2019 | Focusing on ship detection, it is suitable for complex sea surface environment. |
| FUSAR [39] | Gaofen-3 | SAR AIS | 1.124 m × 1.728 m | 512 | 2020 | Focusing on the classification of ships in complex sea surface environment, image slices are stored. |
| - | Sentinel-1 | SAR AIS | Max 4.3 m | – | 2023 | It is used for surface monitoring and is suitable for large-scale scenes such as urban and agricultural. |

### *3.1. Small Target Vessel Identification and Anomaly Analysis and Judgment Framework Based on Multi-Source Information Fusion*

In practical applications, AIS information may be subject to human tampering or shutdown, leading to abnormal AIS information. This often results in mismatches between remote sensing imagery and AIS information. Existing research lacks further discussion and judgment for the matching failures caused by such abnormalities.

Therefore, this paper makes further discussion and judgment on the abnormal situation of matching failure and puts forward the SVIADF, as shown Figure 1.

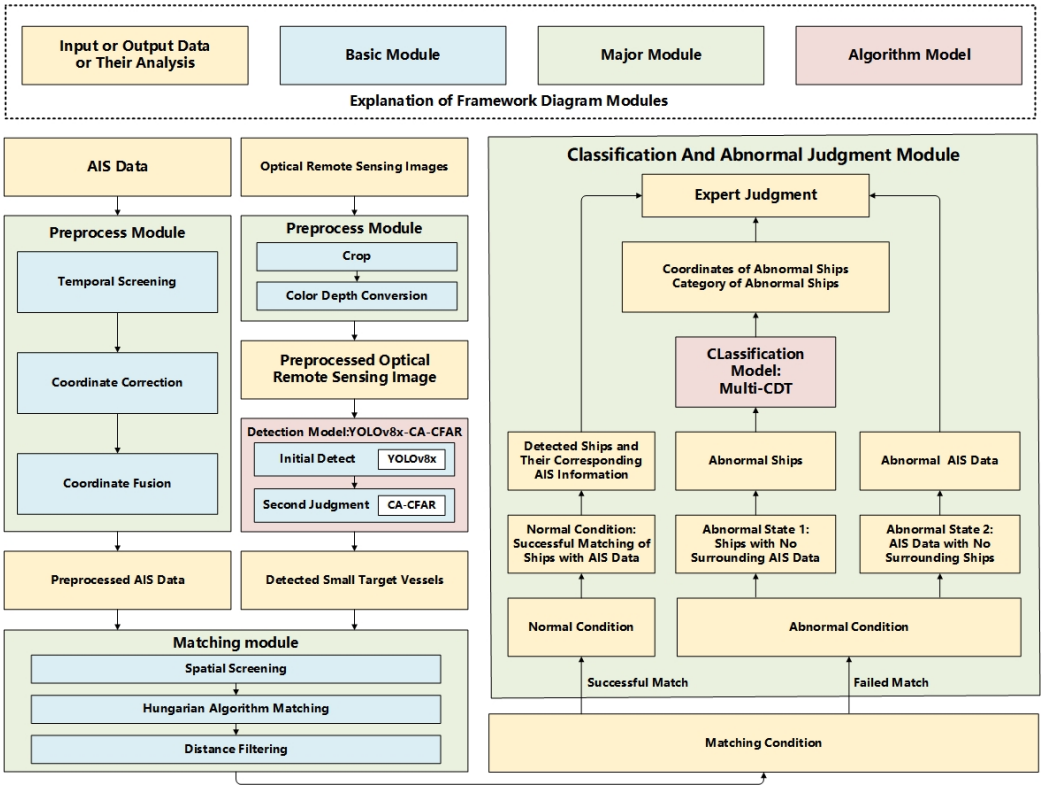



**Figure 1.** Framework of proposed SVIADF method.

Specifically, the input data include AIS information and optical remote sensing images, which need to be preprocessed separately. Then, the small target vessels in the optical remote sensing images are detected, including preliminary detection and secondary judgment, corresponding to the model YOLOv8x-CA-CFAR in Section 3.2. Subsequently, the vessel detection results and the preprocessed AIS information are input into the matching module for matching, followed by an assessment based on the matching outcome.

At this stage, the process can be divided into two broad categories: when a successful match occurs, which is the normal case, the matching results are output. In the case of a match failure, it is considered an abnormal situation, which can be further subdivided into two subcategories: either there is no AIS information around the vessel, or there are no vessels around the AIS information.

When there is no AIS information around the vessel, it is deemed an abnormal vessel. In this case, the detection results are sliced and then input into a classifier trained on AIS–SAR data, corresponding to the model Multi-CDT in Section 3.3, ultimately outputting the coordinates of the abnormal vessel and its predicted category. When there are no vessels around the AIS information, the AIS information is considered abnormal data and is retained. Finally, all the results are saved for expert analysis and judgment.

Due to the different sampling characteristics of remote sensing imagery and AIS information, the former captures static instantaneous vessels positions, while the latter

frequently updates to reflect the dynamic trajectory of vessels. This difference may lead to inconsistent positions of the same vessels in both datasets. In order to match these two kinds of data, based on the empirical advice of cooperating experts, we selected a time window of 3 min above and below the center time of the remote sensing image for AIS data screening instead of directly using the start and end time. The center time is more representative of the representative moment of the image acquisition, reducing the error caused by the delay in the image acquisition process and the target movement. We grouped the filtered AIS data based on MMSI number and then performed the correction and fusion of coordinates respectively. It should be noted that if the duration of AIS data is too long, it may lead to temporal misalignment with remote sensing images and affect the registration accuracy. However, if the duration is too short, it may not cover the motion range of the target, thus affecting the matching effect. The criteria for selecting the data duration are usually based on the moving speed of the target, the temporal resolution of remote sensing image acquisition, and the frequency of AIS data update to ensure that the impact of temporal differences on the matching accuracy is minimized.

Specifically, in order to improve the accuracy in matching, this study preprocessed the AIS data in various ways, including time screening, coordinate correction, and coordinate fusion.

Firstly, time filtering was used to ensure that the selected AIS data matches the time range of image capture, avoiding mismatches due to a large time span. The time filtering equation is formally represented as shown in Equation (1):

$$S_1 = \{x \in \mathrm{S} \mid T_{\text{start}} \leq t(x) \leq T_{\text{end}}\} \tag{1}$$

where $t = \mathrm{df}[\text{updatetime}]$ represents the time update column, $t(x)$ represents the time corresponding to entry $x$ in the dataset, $T_{\text{start}}$ denotes the start time of the image, $T_{\text{end}}$ represents the end time of the image, and $S_1$ represents all entries within the given time interval $[T_{\text{start}}, T_{\text{end}}]$. Based on the CenterTime field information corresponding to each image, $T_{\text{start}}$ and $T_{\text{end}}$ were set respectively for the range of 3 min.

However, some AIS data contain offset parameters, because AIS antennas are usually not installed at the geometric center of the ship. The offset parameter is used to correct this position difference to improve the accuracy of the vessel position reporting. Absence of an offset parameter in the AIS data usually implies that the antenna is installed at the geometric center of the ship, or this information is not provided in the data. It is very meaningful to use the offset parameter to correct the latitude and longitude coordinates in situations where high Precision positioning is required, such as waterway safety, port berthing and collision avoidance. Through this correction, the actual position of the ship can be reflected more accurately and the error can be reduced. Therefore, the offset information in AIS is used for coordinate correction in this study to ensure more accurate position data, and the coordinate correction formula is shown in Equations (2) and (3).

$$Lon\_corrected = Lon + \frac{OffsetA}{k} \tag{2}$$

$$Lat\_corrected = Lat + \frac{OffsetB}{k} \tag{3}$$

where $OffsetA$ represents the east–west directional offset of the ship's position reported by AIS, measured in meters, with positive values indicating the ship is located to the east and negative values indicating the ship is located to the west. Similarly, $OffsetB$ represents the north–south directional offset of the ship's position reported by AIS, measured in meters, with positive values indicating the ship is located to the north and negative values indicating the ship is located to the south. The value of $k$ is an empirical value or calibration

factor that takes into account the magnified meters and converts them into suitable angular units for adjusting the latitude and longitude positions. In this context, $k$ was set to 60,000.

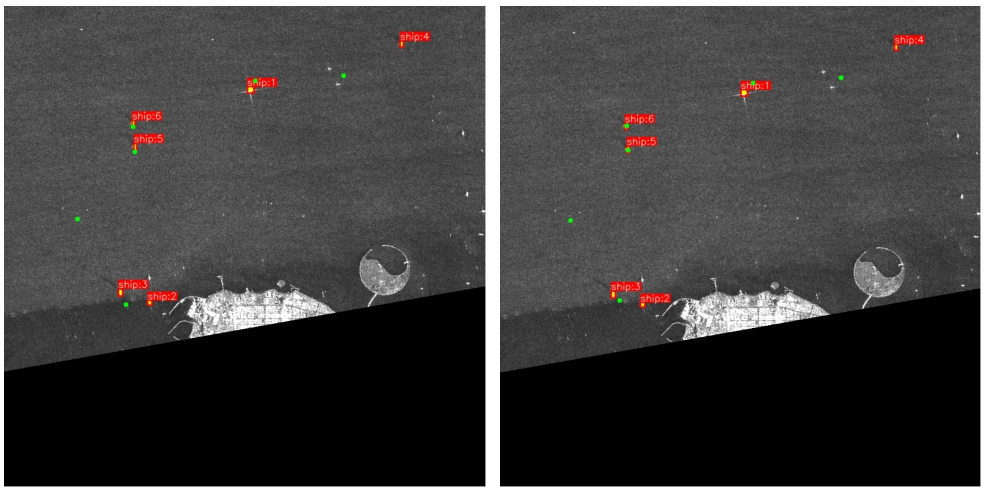
Then, by averaging multiple filtered AIS position points, a more reliable fused coordinate is obtained to reduce the impact of time errors on matching. The coordinate fusion formulas are as follows:

$$Lon_{avg} = \frac{1}{n} \sum_{i=1}^{n} Lon\_corrected_i \tag{4}$$

$$Lat_{avg} = \frac{1}{n} \sum_{i=1}^{n} Lat\_corrected_i \tag{5}$$

where $n = |S_1|$, which represents the number of data entries that meet the conditions after time filtering.

In Figure 2, the fused coordinates before and after rectification are visually evaluated. The yellow points represent the detected positions of vessels, and the green points represent the positions of AIS information. It can be observed that the rectified AIS information positions are closer to the positions of vessels detected on the image than the unrectified positions, indicating that the rectification and fusion preprocessing method is more reliable than simply using the fusion preprocessing method.




(**a**) Original trajectory (**b**) 0.5-min interval

**Figure 2.** The original trajectory and resampled trajectory of the ship.

In summary, after these preprocessing steps, it is possible to match remote sensing imagery and AIS information more accurately, making subsequent analysis more reliable. This processing method not only reduces errors but also addresses challenges such as AIS data anomalies or inconsistencies between imagery and AIS sampling, providing a more robust foundation for vessels detection and classification.

Additionally, conventional remote sensing imagery often has a resolution that is too high for model detection, which requires cropping. Furthermore, some remote sensing images have a 16-bit color depth, differing from the conventional 8-bit color depth, which can lead to display issues. Therefore, processing is required to convert the bit depth.

Before formal matching, spatial filtering based on the geographical scope of the image is also necessary to exclude AIS data outside the image coverage, avoiding the introduction of redundant information and improving matching efficiency. The spatial filtering formula is shown in Equation (6).

$$S_2 = \left\{ x \in S_1 \middle| \begin{array}{l} Lon_{\min} \le Lon_{avg}(x) \le Lon_{\max} \\ Lat_{\min} \le Lat_{avg}(x) \le Lat_{\max} \end{array} \right\} \tag{6}$$

The matching process employs the Hungarian algorithm because it can find the global optimal solution in the bipartite graph matching problem, ensuring the minimization of the total matching distance and avoiding the influence of early decisions on subsequent matching, thereby improving the accuracy and overall efficiency of the matching. Using a greedy algorithm would lead to suboptimal matching, resource wastage, an increase in the rate of error matching, and complex issues in later adjustments.

Given the differences between remote sensing imagery and AIS data, it is not expected that the fused AIS coordinates will match perfectly with the detected vessels coordinates. Therefore, a reasonable matching distance must be set. A recommended maximum tolerance distance of 2 km was used. If the distance between the vessels and the AIS information is within 2 km, the match is considered successful. The mathematical representation of the matching process is given in Equation (7):

$$M = \left\{ (i,j) \in S \times A \middle| \begin{array}{c} \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij} x_{ij} \to \min \\ \text{subject to} \\ \sum_{j=1}^{n} x_{ij} = 1, \quad \forall i \\ \sum_{i=1}^{n} x_{ij} = 1, \quad \forall j \\ x_{ij} \in \{0,1\} \\ d_{ij} x_{ij} < \tau \end{array} \right\} \tag{7}$$

where $M$ represents the set of matching pairs that meet the criteria, which includes effective matches between vessels and AIS information, $S = \{s_1, s_2, \ldots, s_n\}$ represents a set of vessels that need to be matched, $A = \{a_1, a_2, \ldots, a_n\}$ represents the corresponding AIS information, $d_{ij}$ represents the distance between $s_i$ and $a_j$, $x_ij$ represents whether $s_i$ matches with $a_j$, and $\tau$ represents the distance threshold.

### *3.2. Small Target Vessels Detection Based on Constant False Alarm Rate Secondary Judgment*

In the task of small target vessel detection in remote sensing images, research in the SAR (Synthetic Aperture Radar) image domain has combined the CFAR (Constant False Alarm Rate) with deep learning, enhancing the accuracy and applicability of detection algorithms. The CFAR is capable of detecting small-size targets based on the amplitude of pixels, offering its unique advantages.

It is important to note that optical remote sensing imagery and SAR (Synthetic Aperture Radar) remote sensing imagery have many differences in terms of imaging principles, data characteristics, and weather dependence, as shown in Table 5.

**Table 5.** Comparison of differences between optical remote sensing imagery and SAR remote sensing imagery.

| Characteristics | Optical Remote Sensing Imagery | SAR Remote Sensing Imagery |
| --- | --- | --- |
| Imaging Principle | Images are obtained by detecting the reflected light of sunlight or other sources. | Images are constructed by radar waves detecting the ground and receiving the radar signal reflected from the surface. |

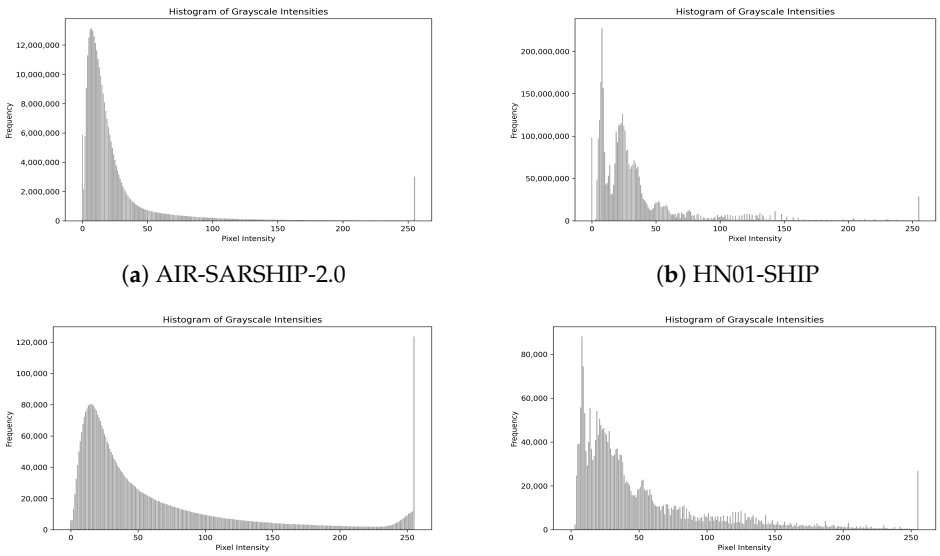
**Table 5.** *Cont.*

| Characteristics | Optical Remote Sensing Imagery | SAR Remote Sensing Imagery |
| --- | --- | --- |
| Data Characteristics | Visual features such as color, texture, and shape are prominent, similar to what the human eye sees. | Represents intensity or phase information, and the image has a certain roughness; not as intuitive as optical images. |
| Weather Dependence | Sensitive to weather conditions; cloud cover and nighttime can affect image acquisition. | Can penetrate clouds and a certain amount of vegetation and can be acquired in all weather conditions and throughout the day. |

Therefore, the CFAR method is not directly applicable to optical remote sensing imagery. So, we extracted grayscale histograms from SAR images and optical remote sensing images to analyze their statistical properties. In Figure 3a,b, the histograms extracted from the SAR image small target ship detection dataset (AIR-SARSHIP-2.0 [55]) and our own optical remote sensing small target ship detection dataset (HN01-SHIP [29]) are presented, respectively. Figure 3c,d display the histograms extracted from the AIR-SARSHIP-2.0 image and the HN01-SHIP image after slice processing. To better analyze the similarities or differences between the histograms, we employed the Bhattacharyya distance, as shown in Equation (8), and the Chi-squared test, as shown in Equation (9), for quantitative comparison.

$$D_B(H_1, H_2) = -\ln\left(\sum_i \sqrt{H_1(i) \cdot H_2(i)}\right) \tag{8}$$

$$\chi^2 = \sum_i \frac{(H_1(i) - H_2(i))^2}{H_1(i)} \tag{9}$$




(**a**) AIR-SARSHIP-2.0 (**b**) HN01-SHIP

(**c**) AIR-SARSHIP-2.0-crop (**d**) HN01-SHIP-crop

**Figure 3.** Comparison of histograms of various data.

Figure 3a with Figure 3b is the first group, and Figure 3c with Figure 3d is the second group. We calculated their Bhattacharyya distance and Chi-squared test values separately, and the results are presented in Table 6.

Based on the Bhattacharyya distance and Chi-squared test values, it is not difficult to observe that the Bhattacharyya distance and Chi-squared test values calculated for the second group, which has undergone slicing processing, are smaller than those for the

first group, which has not been sliced. This indicates that the slicing process effectively reduces the statistical differences between SAR images and optical remote sensing images. Therefore, applying the CFAR to the sliced images is feasible.

**Table 6.** Comparison of similarity metrics.

| Similarity Measure | Group 1 (Not Cropped) | Group 2 (Cropped) |
|---|---|---|
| Bhattacharyya Distance | 0.0003309144084876329 | 0.00012561248225707598 |
| Chi-squared Test Value | 0.0030294396556673595 | 0.0011042250132677447 |

In summary, by reducing domain differences through preprocessing and then applying the concept of transferring domain knowledge, the approach provides a foundation for the use of the CFAR in the field of optical remote sensing images.

In optical remote sensing imagery, a deep learning algorithm can first be used to detect suspicious objects and perform slicing to minimize the differences between SAR and optical remote sensing images. Then, the CA-CFAR can be used on the slices to discriminate, achieving a complementarity in the limitations of deep learning and CFAR and thereby enhancing model performance.

In this study, YOLOv8x was used as the preliminary detection model, and the detection results wer further reviewed to propose the YOLOv8x-CA-CFAR. The pseudocode for the proposed YOLOv8x-CA-CFAR is shown in Algorithm 1.

**Algorithm 1:** YOLOv8x-CA-CFAR algorithm.

```
Input: Image
Output: Detection Results
DetectResults ← YOLOv8xModel(InputImage)
// Step 1: Initial detection using YOLOv8x
if DetectResults.confidence > 0.05 then
    if RetainedResults.confidence > 0.25 then
        RetainedResults ← DetectResults
    else if DetectResults.isSuspicious then
        // Perform CA-CFAR refinement
        Thresholds ← CA-CFAR(img, gSize, tSize, pfa)
        Regions ← applyThresh(img, Thresholds)
        // Check if Regions are not empty
        if Regions is not empty then
            RetainedResults ← DetectResults
        else
            // Discard the result
            RetainedResults ← null
    else
        // Discard the result
        RetainedResults ← null
else
    // Discard the result
    RetainedResults ← null
return RetainedResults
```

The original YOLOv8x algorithm only retains detection boxes with confidence scores above 0.25, which reduces false positives but may also miss small vessels with less-prominent features, affecting Recall rate and detection accuracy.

To address this issue, the YOLOv8x-CA-CFAR algorithm temporarily retains detection boxes with confidence scores between 0.05 and 0.25 and further analyzes these suspicious areas using the CA-CFAR algorithm. If the CA-CFAR confirms the presence of a vessel, and the detection box is retained; otherwise, it is discarded.

This method improves the ability to recognize small vessels while maintaining high detection accuracy, and the final result integrates detection boxes with high confidence scores and those confirmed by CA-CFAR, achieving a more comprehensive and accurate detection output.

### *3.3. Unsupervised Domain-Adaptive Small Target Vessel Classification Based on Multi-Source Heterogeneous Data*

The existing unsupervised domain-adaptive transfer learning research mainly focuses on the homogeneous or heterogeneous transfer of a single-source domain, and even if a multi-source domain is involved, most of them are limited to the transfer of multi-source homogeneous data. In [61], the fusion use of heterogeneous data in transform detection is mentioned, but it refers to optical data with different resolutions. In [62], although unsupervised domain adaptation of multi-source heterogeneous data is also mentioned, it refers to the difference in the feature dimension of the source domain. The concept of multi-source heterogeneous data mentioned in this paper is different from the former.

In Figure 4, the visual representation of several cases mentioned in this paper is given. In general, there is a lack of research in the field of unsupervised domain adaptation of multi-source heterogeneous data.

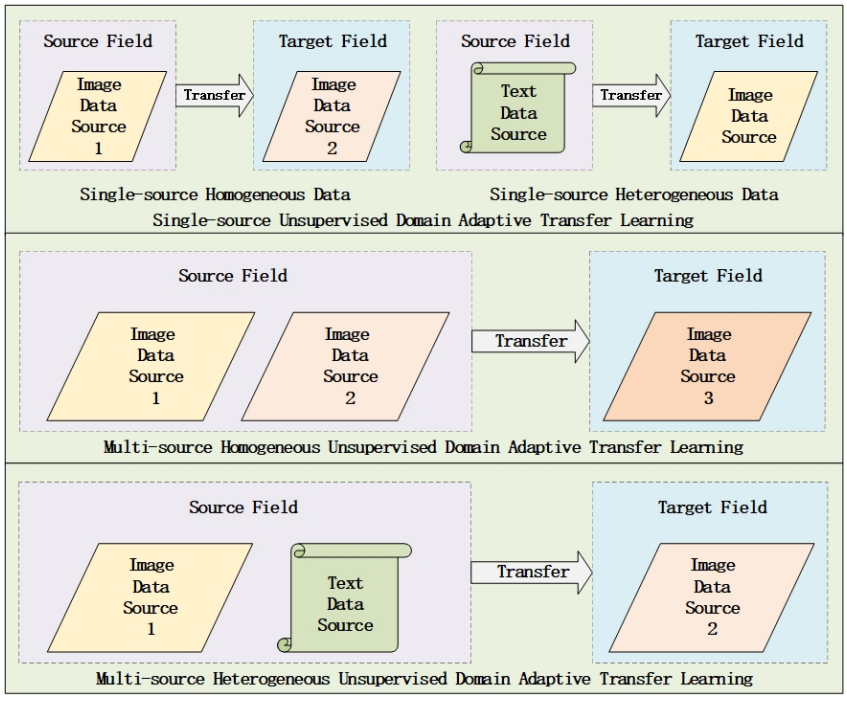



**Figure 4.** Comparison of unsupervised domain adaptation scenarios.

In this paper, we transfered the heterogeneous sources of SAR imagery and AIS information to optical remote sensing imagery. We propose a multi-source heterogeneous unsupervised domain adaptation model called Multi-CDT based on Transformer. The model architecture is depicted in Figure 5.

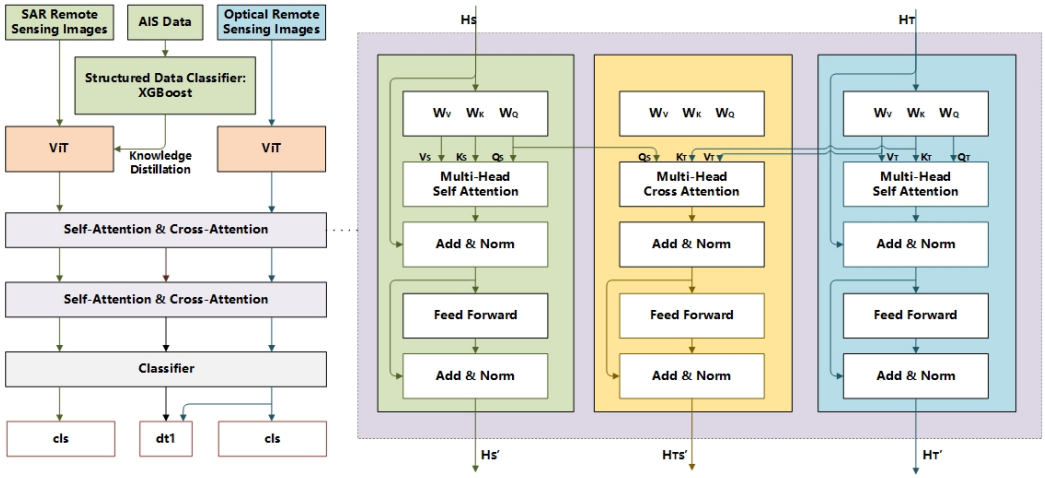



**Figure 5.** Multi-CDT network architecture.

Specifically, considering the differences in imaging principles among different remote sensing images, which can lead to variations in pixel distribution, the source domain data used in this paper were the FUSAR dataset [39]. Each SAR remote sensing image has corresponding AIS information, which includes the length and width information of the vessels on each image.

In [40], based on the length and width information, 16 types of features were designed, such as perimeter, area, aspect ratio, and other geometric features, the details are shown in Table 7.

**Table 7.** Calculation methods of 16 characteristics based on ship length and width design.

| Characteristic Symbol (Meaning) | A Formula or Source of Calculation |
|---|---|
| F1 (Length) | The Length of Ship |
| F2 (Width) | The Width of Ship |
| F3 (Perimeter) | $2 \times (\text{Length} + \text{Width})$ |
| F4 (Area) | $\text{Length} \times \text{Width}$ |
| F5 (Aspect Ratio 1) | $\text{Length}/\text{Width}$ |
| F6 (Aspect Ratio 2) | $\text{Width}/\text{Length}$ |
| F7 (Shape Complex) | $(\text{Length} + \text{Width})^2/(\text{Length} \times \text{Width})$ |
| F8 | $\text{Length} - \text{Width}$ |
| F9 | $\text{Length}/(\text{Length} + \text{Width})$ |
| F10 | $\text{Width}/(\text{Length} + \text{Width})$ |
| F11 | $(\text{Length} - \text{Width})/(\text{Length} + \text{Width})$ |
| F12 | $(\text{Length} - \text{Width}) \times (\text{Length} + \text{Width})$ |
| F13 | $\text{Length}^2/(\text{Length}^2 + \text{Width}^2)$ |
| F14 | $\text{Width}^2/(\text{Length}^2 + \text{Width}^2)$ |
| F15 | $(\text{Length} \times \text{Width})/(\text{Length}^2 + \text{Width}^2)$ |
| F16 | $(\text{Length}^2 - \text{Width}^2)/(\text{Length}^2 + \text{Width}^2)$ |

Considering that these geometric characteristics are not easy to change with the different imaging principles of the image, they were used in Multi-CDT to train an additional classifier based on AIS features.

The AIS-based classifier was then used as a teacher model to guide the distillation learning during the pretraining of ViT. This process extracts features that are more discriminative and robust, improving the performance of the ViT model while also enhancing the performance of the model that is later transferred to optical remote sensing images.

During the pretraining phase, the loss function is divided into two parts: the first part is the classification loss, as shown in Equation (10), and the second part is the distillation loss, as shown in Equation (12).

$$L_{cls} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C} y'_{i,j}\log\left(\frac{e^{x_{i,j}}}{\sum_{k=1}^{C} e^{x_{i,k}}}\right) \tag{10}$$

where $N$ is the number of samples in a batch, $C$ is the number of classes, $x$ is the logits output of the model (before applying the softmax function), $\mathbf{x}_{i,j}$ represents the model's output for the $i$-th sample corresponding to the $j$-th class, $log$ denotes the natural logarithm, $e$ is the base of the natural logarithm used for calculating the exponent, and $\mathbf{y}'_{i,j}$ represents the smoothed label, as shown in Equation (11).

$$y'_{i,j} = (1-e)\cdot y_{i,j} + \frac{e}{C} \tag{11}$$

where $y$ is the one-hot encoded true label. The label smoothing parameter is $e$.

By reducing the model's reliance on accurate labels and increasing its small probability predictions for all categories, the model becomes more robust when facing uncertainty, improving its generalization ability.

Common multi-source information fusion often requires alignment at the feature level. However, the data types, distributions, and feature expression methods between different data sources may vary greatly, such as text, image data, etc. This heterogeneity can lead to information loss or inaccurate alignment during alignment, especially when there is a lack of explicit association between data modalities. Through the distillation loss shown in Equation (12), knowledge transfer is achieved at the decision level.

$$L = \frac{1}{N}\sum_{i=1}^{N}\left(-\sum_{k=1}^{K} Ais\_score_{ik}\cdot\log(\text{Image}\_score_{ik})\right) \tag{12}$$

where $N$ is the number of samples (i.e., batch size), $K$ is the number of classes, $Ais\_score_{ik}$ represents the probability of the $i$-th sample in the $k$-th category given by the AIS information classifier, and Image_$score_{ik}$ represents the probability score of the $i$-th sample in the $k$-th category given by the image classifier.

Considering that the geometric features extracted from AIS information are structured data, XGBoost [63] was used as the classifier model based on AIS features in the Multi-CDT. Here, other classifier models could also be used.

## 4. Experimental Results and Analysis

### *4.1. Small Target Vessel Detection Task*

#### 4.1.1. Experimental Environment and Parameter Settings

The training and testing environment for the experiments was the Windows 10 operating system, a NVIDIA TITAN RTX GPU, PyTorch 1.9, and Python 3.6. In the training process, the number of epochs was set to 300, with each epoch iterating 71 times. The batch size was 64, and the initial learning rate was 0.01.

#### 4.1.2. Dataset Introduction

We have developed a dataset specifically designed for detecting small target vessels in high-resolution, wide-area optical remote sensing imagery [29]. Specifically, the dataset was constructed based on remote sensing images from the Hainan Satellite 1.

Due to the large resolution of the original images, each image was first split into smaller parts. Data augmentation was then performed using three different image processing techniques. Subsequently, the images were annotated using LabelMe.

Ultimately, the dataset consists of 3829 images with a resolution of 1024 × 1024. The original image dataset was then randomly divided into a training set (containing 2679 images), a test set (containing 345 images), and a validation set (containing 805 images) in a specific proportion using a Python script.

#### 4.1.3. Performance Indicators

(1) Precision

Precision measures the accuracy of the model's predictions for positive samples. Its calculation is as shown in Equation (13), representing the proportion of positive samples that the model correctly predicted as positive.

In the context of object detection tasks, Precision refers to the proportion of actual small targets that are detected, reflecting the situation of false detections.

$$Precision = \frac{TP}{TP + FP} \tag{13}$$

(2) Recall

Recall measures the completeness of the model's predictions for actual positive samples. Its calculation is as shown in (14), representing the proportion of positive samples that the model correctly predicted as positive relative to the total number of actual positive samples.

In object detection tasks, Recall refers to the proportion of actual small targets that are detected, reflecting the model's missing detection of small targets.

$$Recall = \frac{TP}{TP + FN} \tag{14}$$

(3) F1-score

The F1-score is a composite measure that takes into account both Precision and Recall. It is the harmonic mean of these two metrics. Since Precision and Recall often trade off against each other, the F1-score is used to balance them, aiming to find the best model performance that lies between the two. The F1-score ranges from 0 to 1, with higher values indicating better model performance. When both Precision and Recall are high, the F1-score will also be correspondingly high.

$$F_1 = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{15}$$

#### 4.1.4. Results and Analysis

In marine surveillance tasks, especially for small vessels, the importance of the F1-score increases significantly because this metric takes into account the Precision and Recall of the model to ensure that any potential illegal activity or security threat is identified to the greatest extent possible while reducing false positives. This balance is particularly critical because of the complex marine environment and the small size of the target, which is easy to miss in the vast ocean. A high F1-score means that the model can effectively capture key targets, reduce missed detections, and control false positives, ensuring the accuracy and timeliness of marine safety monitoring, thus supporting effective law enforcement and resource management decisions.

In the experimental comparison, in order to better reflect the influence of YOLOv8x-CA-CFAR on the second judgment of the low-confidence detection results using CA-CFAR, YOLOv8x-LT was further added to the experiment (simply reducing the threshold conf = 0.05. That is, all detection results with a confidence greater than 0.05 are retained) for comparative analysis. In theory, when the confidence threshold is reduced, although more detection results with low confidence can be identified as ships and the Recall rate of the model can be directly improved, it will also lead to many misjudgments and reduce the accuracy. Testing the images on the test set, the performance of various networks is shown in Table 8.

**Table 8.** Performance comparison of various networks.

| Networks | Precision (%) | Recall (%) | F1-Score (%) |
|---|---|---|---|
| Faster R-CNN [64] | 72.3 | 68.4 | 70.3 |
| SSD [65] | 71.8 | 67.7 | 69.7 |
| YOLOv5n [66] | 86.5 | 83.5 | 85.0 |
| YOLOv7 [67] | 87.5 | 84.5 | 86.0 |
| YOLOv7x [68] | 88.2 | 85.1 | 86.6 |
| YOLOv7&QFL [69] | 88.6 | 86.7 | 87.6 |
| YOLOv8n [34] | 89.1 | 87.6 | 88.3 |
| YOLOv8x [34] | 94.0 | 95.3 | 94.6 |
| YOLOv8x-LT | 90.7 | 98.8 | 94.6 |
| YOLOv8x-CA-CFAR | 93.5 | 98.2 | 95.8 |

In the experiments, the performance of different YOLO models varied, and the YOLOv8 variant performed particularly well. YOLOv8n achieved a high F1-score (88.3%) with its lightweight design, which is suitable for real-time applications with limited resources. For the original YOLOv8x (conf = 0.25), the Precision, Recall, and F1-score of YOLOv8x-LT (simply reducing the threshold conf = 0.05) were 94.0%, 95.3%, and 94.6%, respectively. At 98.8% and 94.6%, the Precision was reduced by 3.3%, the Recall rate was increased by 3.5%, and the F1-score remained unchanged. YOLOv8x-CA-CFAR performed a second judgment on the suspicious targets with confidence values between 0.05 and 0.25, screened the detection results in this interval, and finally retained the filtered results and all the detection results with confidence values greater than 0.25 as the final detection results. Its Precision, Recall, and F1-score were 93.5%, 98.2%, and 95.8%, respectively. The Precision was only reduced by 0.5%, the Recall was increased by 2.9%, and the F1-score was increased by 1.2%.

The performance improvement is attributed to YOLOv8x-CA-CFAR's combination of the powerful feature extraction capability of deep learning algorithms and the advantages of CA-CFAR in small target detection. By employing secondary discrimination, it improves detection Precision and Recall, compensating for the shortcomings of the simple threshold reduction method and thereby significantly enhancing overall performance. The CA-CFAR method adaptively processes the background noise in the detection area, dynamically adjusting the detection threshold to effectively filter out background noise and false positives and maintaining high Precision even at a low confidence threshold, thus achieving the best balance between detection comprehensiveness and accuracy.

In addition, Figures 6 and 7 show the detection results of YOLOv8x, YOLOv8x-LT, and YOLOv8x-CA-CFAR, where the green box and green "GT" are the real sample labels, and the red box represents the detection results retained by the original YOLOv8x. That is, the detection results with the confidence values of the predicted box greater than 0.25, and then the yellow box and yellow "0.05–0.25" are used to mark the additional detection

results retained by YOLOv8x-LT in addition to the detection results of YOLOv8x by simply reducing the threshold. The purple box and purple "ca-cfar" mark the results of YOLOv8x-CA-CFAR in addition to the detection results of YOLOv8x, and the suspicious objects with a confidence interval of 0.05–0.25 were also retained after the second judgment of CA-CFAR.

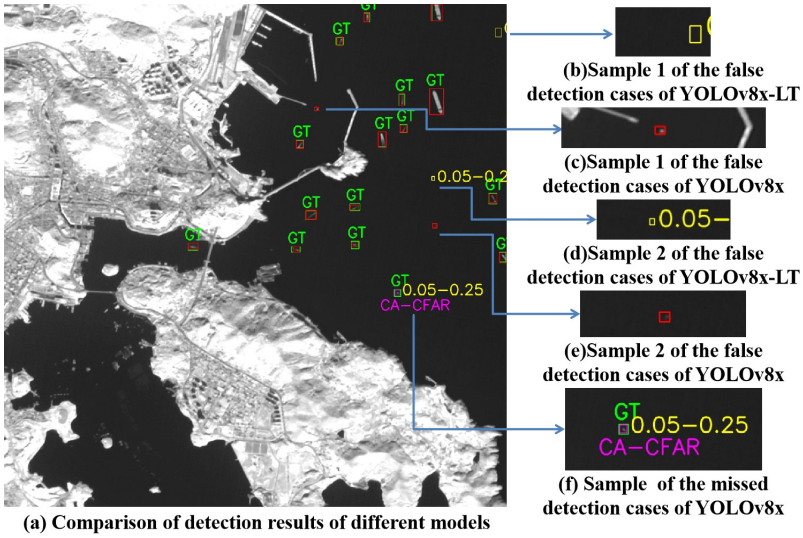



**Figure 6.** Example 1 of the comparison of the detection results of the three models.

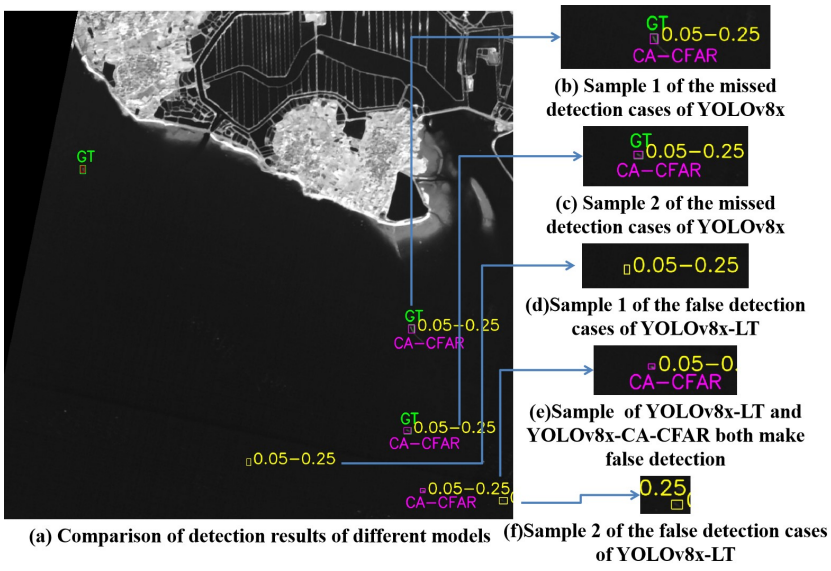



**Figure 7.** Example 2 of the comparison of the detection results of the three models.

From the comparison of the detection results of the models in Figures 6 and 7, it can be seen that although the prediction results of the original YOLOv8x model are generally consistent with the real labeling box, there were some false detections (see Figure 6c,e) or missed detections (see Figures 6f and 7b,c). In engineering applications, we pay more attention to the Recall rate of the model, and the risk of missing detection may be greater when monitoring illegal activities and potential threats. By increasing the Recall, we can ensure that more potential targets are detected, even though these targets may not exactly meet the high standard accuracy of the model. From Figures 6f and 7b,c, it can be seen that YOLOv8x-LT and YOLOv8x-CA-CFAR could retain the missing small target ship detection results, but from the other detection results, it can be found that although YOLOv8x-LT improved the Recall rate, However, a large number of false positives were generated (see Figures 6b,d and 7d–f), and if too many false positives are generated, it may occupy a lot of

resources for unnecessary inspection and follow up, affecting the efficiency of maritime safety operations. Therefore, a more refined approach is needed to balance Recall and Precision, and the proposed YOLOv8x-CA-CFAR reduces the number of missed detections compared to YOLOv8x, which detects small-size ships, although it also generates a false positive in the example (Figure 7e). However, compared to YOLOv8x-LT, the number of false detections was significantly reduced.

In addition, the performance quantitative comparison in the two comparison examples is given in Table 9.

**Table 9.** Quantitative performance comparison of two comparison examples.

| Example | Networks | Precision (%) | Recall (%) | F1-Score (%) |
|---|---|---|---|---|
| Example 1 | YOLOv8x | 88.24 | 93.75 | 90.91 |
| | YOLOv8x-LT | 80.00 | 100.00 | 88.89 |
| | YOLOv8x-CA-CFAR | 88.89 | 100.00 | 94.12 |
| Example 2 | YOLOv8x | 100.00 | 33.33 | 50.00 |
| | YOLOv8x-LT | 50.00 | 100.00 | 66.67 |
| | YOLOv8x-CA-CFAR | 75.00 | 100.00 | 85.71 |

From the specific quantitative results of the two comparison examples in Table 9, it can also be seen that YOLOv8x-CA-CFAR greatly improved the Recall rate and the comprehensive performance of the model at the expense of a small accuracy and achieved the optimal F1-score.

Finally, Figure 8 further shows more detection results of the YOLOv8x, YOLOv8x-LT, and YOLOv8x-CA-CFAR models, where the missed detection results are marked by red circles, and the false detection results are marked by yellow circles. It is easy to see that compared to other methods, YOLOv8x-CA-CFAR achieved an effective balance between Recall and Precision and significantly reduced the incidence of false positives while ensuring a high Recall rate, which effectively optimizes the overall performance of maritime safety monitoring.

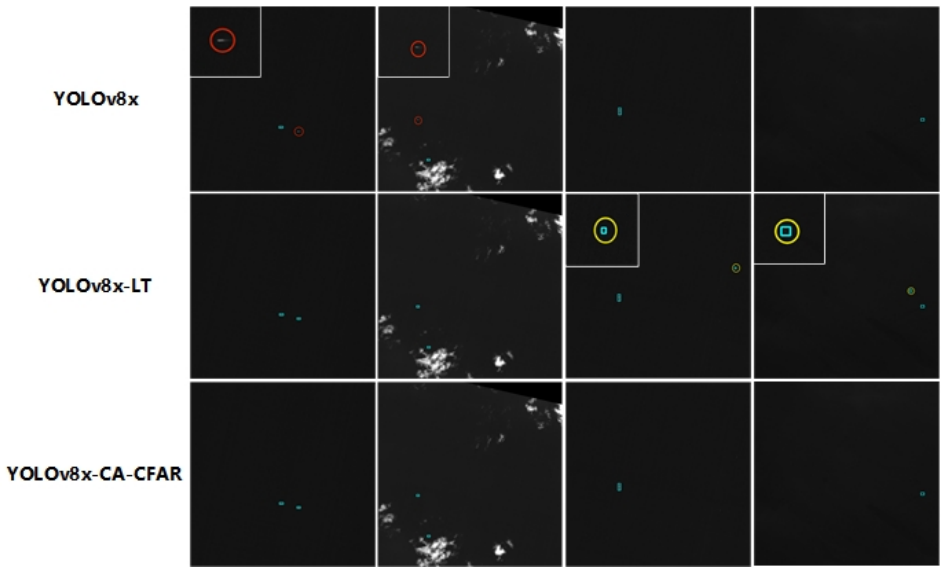



**Figure 8.** Comparison of the detection results of the three models.

In addition, to better simulate real-world application scenarios, we conducted tests on the entire visualization-based multi-source information fusion small target vessel recognition and anomaly judgment software. By inferring, matching, displaying, and saving

correspondences on nearly 800 images, we obtained an average running time of only 0.083 s per image, which is still within the acceptable range of real-time performance. The test results show that in the scenario of large-scale data processing, the proposed model can not only ensure high Recall rate and Precision but also maintain efficient processing speed to ensure efficient performance in complex real-time marine safety monitoring tasks.

### *4.2. Small Target Vessel Classification Task*

#### 4.2.1. Experimental Environment and Parameter Settings

The training and testing environment for the experiment was a Windows 10 operating system equipped with a NVIDIA TITAN RTX GPU, PyTorch 1.9, and Python 3.6. In the training process, the pretrained epochs were set to 50, and the transfer training epochs were also set to 50. Both the pretraining and transfer training used a batch size of 64 and an initial learning rate of 0.01.

#### 4.2.2. Dataset Introduction

The dataset was constructed based on the FUSAR [39] public dataset and remote sensing images from the Hainan Satellite 1. The FUSAR dataset contains ship images with a resolution of 512 × 512, divided into 15 major categories, with a total of 5244 images. To construct the optical remote sensing image classification dataset (target domain data), the detection dataset from the Hainan Satellite 1 was processed. The images of the vessels were sliced using annotation boxes. To prevent distortion of the vessel proportions during the subsequent resizing process, the rectangular images obtained from the slicing were filled with black pixels to make them square, thereby maintaining classification performance. Finally, 3814 small target ship classification images based on the Hainan Satellite 1 were obtained as the unlabeled target domain dataset.

#### 4.2.3. Performance Indicators

In practical engineering, it is difficult to obtain classification labels for high-resolution wide-area optical remote sensing images. Quantitative analysis mainly relies on source domain labels and pseudolabels for the target domain. This section of the experiment focuses on qualitative analysis, and the following techniques were used in the qualitative analysis experiment:

(1) Grad-CAM (Gradient-weighted Class Activation Mapping)

Grad-CAM [70] is a technique for enhancing the visualization of convolutional neural networks (CNNs) used in image recognition and classification tasks. It demonstrates the relationship between the prediction of a specific category and the activation strength at different positions in the network by using the gradients of the last convolutional layer for the target category. This helps researchers understand which parts of the image the network focuses on for classification.

(2) T-SNE (T-distributed Stochastic Neighbor Embedding)

T-SNE [69] is a technique for dimensionality reduction and visualization of high-dimensional data. It maps high-dimensional data points to a low-dimensional space (usually two or three dimensions) for visualization while preserving the local structure between data points, i.e., similar data points remain close in the low-dimensional space.

#### 4.2.4. Results and Analysis

In CDTrans, unsupervised domain adaptation learning was divided into two steps: first, ViT was used for pretraining on the source domain, and then the trained ViT model was used for transfer learning. In the quantitative experiment, CDTrans and Multi-CDT

were trained for 50 epochs on both the source and target domains, while other models were trained for 100 epochs to ensure fair comparison.

The experimental results, as shown in Table 10, demonstrate that Multi-CDT exhibited excellent classification ability on both the source and target domains with an accuracy of 99.82%, significantly outperforming other transfer learning methods. This indicates that Multi-CDT, through multi-source information fusion, extracts features that are more discriminative and robust, making it suitable for complex and changing real-world scenarios.

**Table 10.** Performance comparison of various models on the classification task.

| Model | Source Domain (True Labels) | Target Domain (Pseudo Labels) |
|---|---|---|
| DAN [41] | 92.00% | - |
| JAN [43] | 94.40% | - |
| DANN [47,48] | 96.70% | - |
| AFN [49] | 98.30% | - |
| MCD [51] | 98.30% | - |
| BSP [52] | 93.20% | - |
| CDT [53] | 99.80% | 99.50% |
| Multi-CDT | 99.82% | 99.82% |

Additionally, Figure 9 and Figure 10 show the training trends of Multi-CDT and CDT in the source domain pretraining task and the target domain unsupervised domain adaptation training task, respectively.

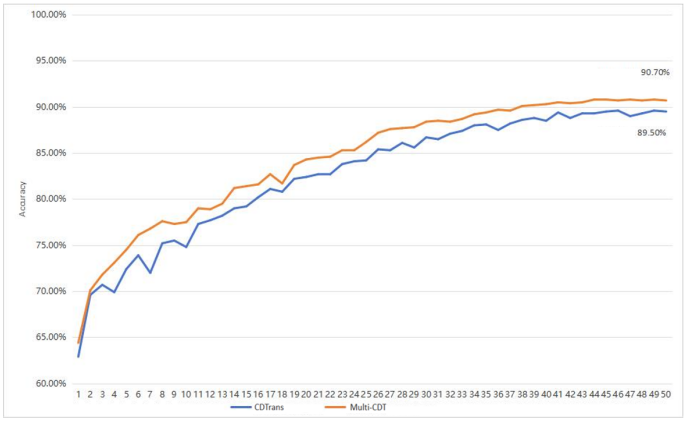



**Figure 9.** Source domain pretraining trend.

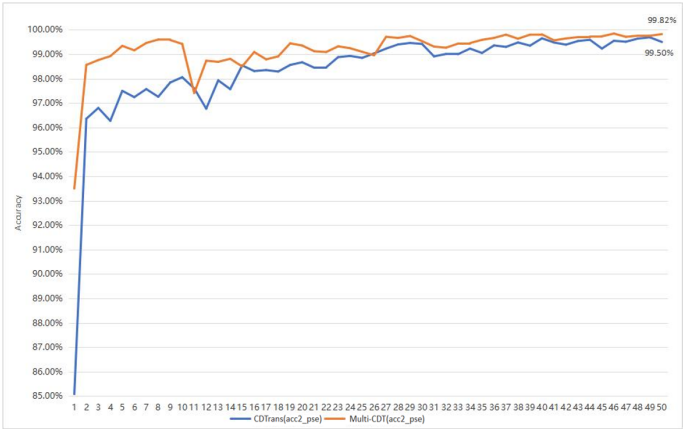



**Figure 10.** Target domain training trend.

As shown in Figure 9, the Multi-CDT model demonstrated higher accuracy and faster improvement on the test set, indicating that it captures the data correlation and features better through multi-source information fusion, thereby enhancing performance. In contrast, the accuracy improvement of the CDTrans model was slower, and its final performance is inferior to that of the Multi-CDT model. At the end of pretraining, Multi-CDT had a 1.2% higher accuracy than CDTrans.

Figure 10 illustrates the performance comparison of Multi-CDT and CDTrans in the unsupervised transfer learning task. The Multi-CDT model showed higher accuracy from the early stages of the experiment and maintained an accuracy of over 98% throughout the training process, demonstrating stable and consistent high performance. In contrast, the accuracy of the CDTrans model was initially low but gradually improved over time, stabilizing around the 20th epoch. The accuracy of Multi-CDT was close to 100% with minimal fluctuation, while CDTrans achieved a maximum accuracy of 99.5% but started with an accuracy below 90%. Overall, Multi-CDT exhibited better initial performance and stability, and its multi-source information fusion strategy enabled it to utilize source domain information more effectively to enhance target domain performance.

Figure 11 and Figure 12 illustrate the heat maps produced by the models trained with CDTrans and Multi-CDT during the second stage of domain adaptation applied to the source domain image (SAR remote sensing image) and the target domain image (optical remote sensing image), respectively. In the heatmap, reddened areas indicate that the model is more focused on this area of the image. Furthermore, these two models were utilized to extract features from the target domain image (optical remote sensing image). A T-SNE plot was subsequently generated, as depicted in Figure 13. In the following, we will carry out qualitative analysis based on these figures.

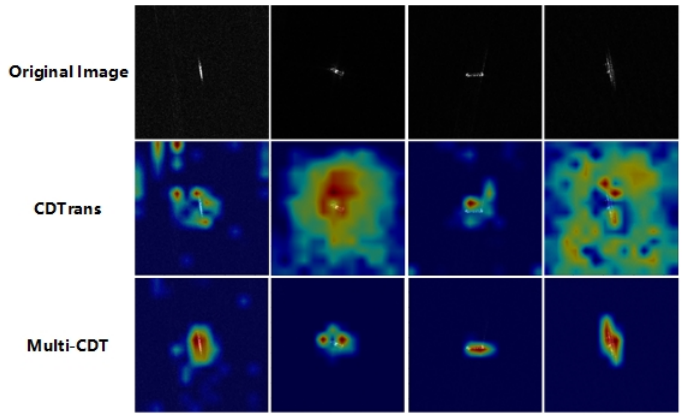



**Figure 11.** Attention map based on source domain image (SAR remote sensing image).

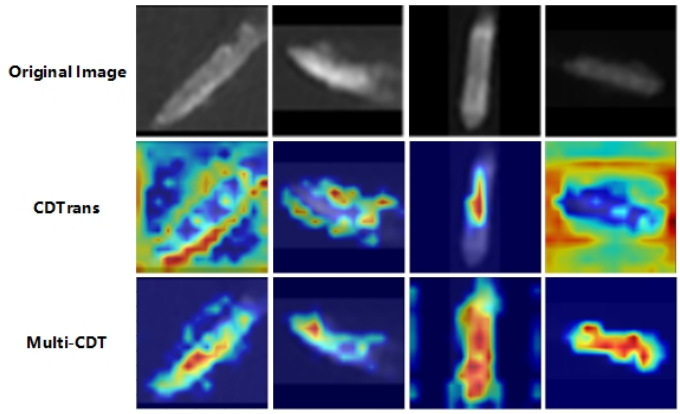
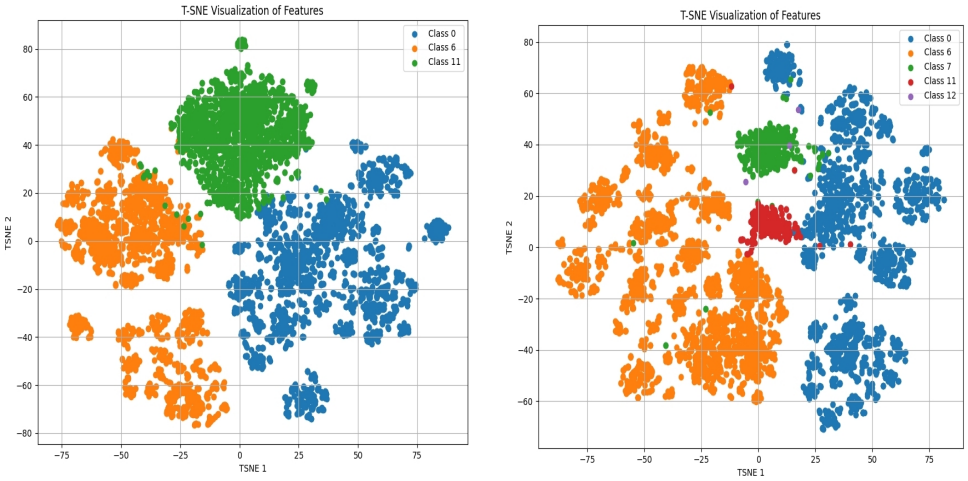



**Figure 12.** Attention map based on target domain image (optical remote sensing).

According to the attention maps shown in Figures 11 and 12, it can be observed that the proposed Multi-CDT method could capture important areas more accurately than CDTrans and pay less attention to the background, with more hotspots on the target objects.




(**a**) CDTrans (**b**) Multi-CDT

**Figure 13.** T-SNE plot based on target domain image (optical remote sensing image).

In practical classification tasks, it is important for the model to accurately distinguish the diverse categories of the source domain rather than simplifying them into just a few large categories. Considering there are 15 categories in total, including one representing "other" as class6, a further detailed analysis of the category distribution was conducted through two T-SNE plots.

Figure 13a shows three main categories (blue, orange, green), with the orange category densely clustered in the lower half of the plot, suggesting that the model may have grouped multiple categories into "other". The blue and green categories show good clustering but are insufficient to reflect the diversity of categories in the source domain.

Figure 13b displays five main categories (blue, orange, green, red, and purple), providing a more detailed categorization and demonstrating the model's better ability to recognize different categories. The orange category is more widely distributed, and the boundaries between the blue, green, and red categories are clearer. Although there is some overlap and mixing, the category distribution is richer, better reflecting the diversity of categories in the source domain.

### *4.3. Small Target Vessel Anomaly Detection Task*

#### 4.3.1. Experimental Environment and Parameter Settings

The training and testing environment for the demonstration was a Windows 10 operating system equipped with a NVIDIA TITAN RTX GPU, PyTorch 1.9, and Python 3.6.

#### 4.3.2. Dataset Introduction

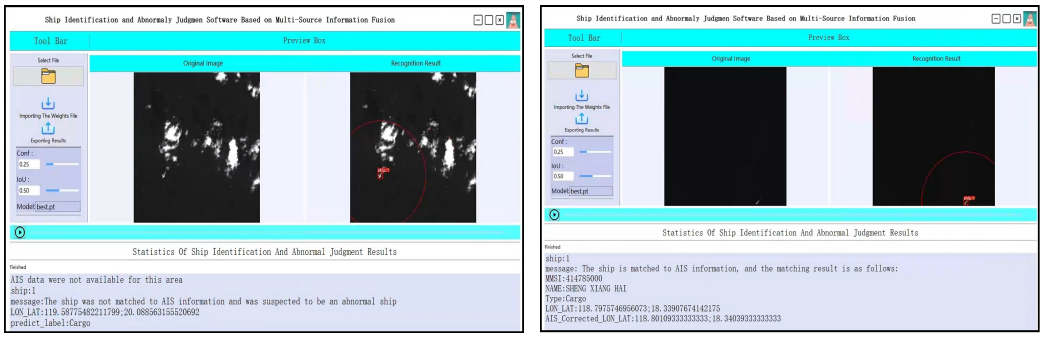
In this study, we first used the optical remote sensing images and AIS data obtained by Hainan Satellite 1 for demonstration, and it is also worth mentioning that the method proposed in this study is not only applicable to the Hainan Satellite 1 data but also has good data source adaptation ability. Therefore, in order to further verify the versatility and adaptability of this method, we also used Sentinel-1 data for supplementary verification.

#### 4.3.3. Real-World Demonstration

In practical applications, we label and index the detected vessels and draw a circular area with a radius of 2 km centered on each vessel to indicate the visible matching range.

If the vessels within this range can be matched with AIS data, the system will mark them with green dots; if not, they will be marked with red dots.

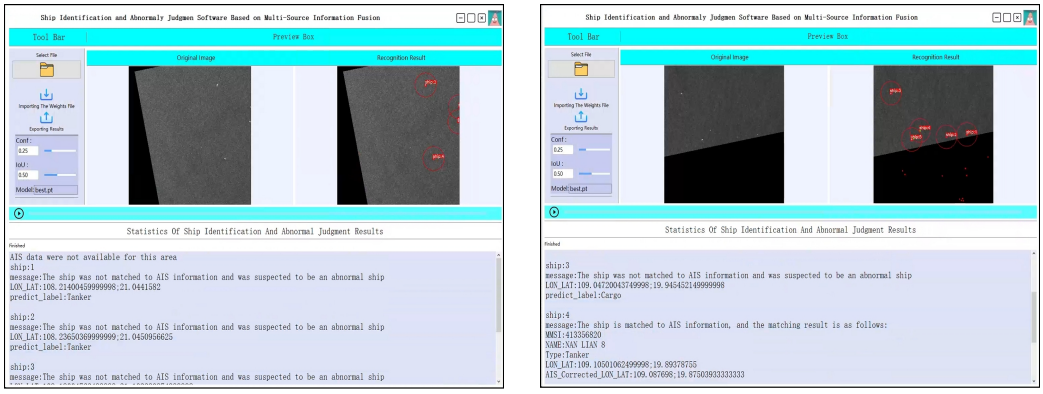
For image regions where AIS data are unavailable, the system issues a warning with the message, "AIS data were not available for this area". It simultaneously highlights the locations of anomalous vessels and displays the classifier's prediction results, as illustrated in Figure 14a. When AIS data are available for the image region, the system automatically attempts to perform matching and displays the results alongside the detection overview. An example of this scenario is shown in Figure 14b. This visualization approach not only provides a clear and intuitive representation of the detection and matching process but also facilitates timely anomaly identification and decision making in maritime monitoring and management.




(**a**) AIS data is not available (**b**) AIS data is available

**Figure 14.** Demonstration example based on Hainan Satellite 1 data.

To further validate the applicability of the proposed method, we conducted demonstrations using Sentinel-1 data. Figure 15a and Figure 15b illustrate scenarios without and with AIS data, respectively. These results demonstrate that the proposed approach effectively performs anomaly detection across different data sources, maintaining its reliability and adaptability.




(**a**) AIS data is not available (**b**) AIS data is available

**Figure 15.** Demonstration example based on Sentinel-1 data.

In summary, the practical tests confirm that the multi-source information fusion anomaly detection framework proposed in this study exhibits strong robustness. It consistently delivers accurate and reliable detection results across diverse real-world application scenarios. This robustness underscores its potential as a valuable tool for maritime monitoring, enhancing the capability to address challenges posed by varying data availability and environmental conditions.

In addition, in order to be closer to the actual application scenario, we tested the whole visual multi-source information fusion small target ship recognition and anomaly

judgment software. By inferring, matching, displaying, and saving correspondences on nearly 800 images, we obtained an average running time of only 0.083 s per image, which is still within the acceptable range of real-time performance. The test results show that in the scenario of large-scale data processing, the proposed model can not only ensure high Recall rate and accuracy but also maintain efficient processing speed to ensure efficient performance in complex real-time marine safety monitoring tasks.

### *4.4. Discussion and Future Works*

#### 4.4.1. Small Target Vessel Detection Task

This study addresses the issue of fewer pixels and less-distinct features of ships in high-resolution wide-area optical remote sensing imagery by proposing a model that combines YOLOv8x with CA-CFAR—named YOLOv8x-CA-CFAR. The model integrates the traditional CA-CFAR algorithm with the advanced YOLOv8x deep learning model, achieving a fusion of detection methods based on structure and pixels, and explores a new path for improvement.

The CA-CFAR algorithm has significant advantages in the detection of small targets in noisy environments, especially in the field of SAR images. By adaptively adjusting the local background noise, it enables more flexible threshold setting, which is highly suitable for the detection of small targets in remote sensing images. However, there is a significant theoretical difference between optical remote sensing images and SAR remote sensing images, making it impossible to directly apply CFAR to optical remote sensing images. This study, however, uses a deep learning model for preliminary detection and performs image slicing based on the detection results, reducing the differences between images and allowing the advantages of CFAR to be leveraged.

The YOLOv8 series models, with their powerful feature extraction capability, can effectively detect suspicious targets in complex backgrounds. Experimental results show that the secondary judgment strategy in the proposed YOLOv8x-CA-CFAR model effectively avoids the problem of increased false positives caused by simply reducing the confidence threshold. Compared to the YOLOv8x-LT model, which only improves Recall rate by adjusting the threshold, YOLOv8x-CA-CFAR, by utilizing the adaptive background noise suppression characteristic of CFAR, shows a significant advantage in reducing false positives, thereby further increasing the F1-score, validating the feasibility and forward-looking nature of the "traditional + intelligent" fusion method. This also illustrates the generality of knowledge across different domains.

In future research, considering that CA-CFAR still belongs to the traditional CFAR algorithm, we can try to combine the more advanced CFAR algorithm with deep learning. Simultaneously, research can explore additional preprocessing methods to further mitigate the differences between various remote sensing domains. In addition, it will try to calculate the length and width of the detected ships based on the detection results so as to provide more abundant information for ocean detection. Finally, we will try to further improve the reliability of small target ship detection based on multi-source data fusion.

#### 4.4.2. Small Target Vessel Classification Task

To address the challenge of obtaining classification sample data from high-resolution wide-area optical remote sensing imagery, this paper proposed a solution using unsupervised domain adaptation. Existing research on unsupervised domain adaptation transfer learning primarily focuses on the transfer of homogeneous or heterogeneous data from a single-source domain. Even when dealing with multiple source domains, most research is limited to the transfer of homogeneous data across these domains. However, in practical applications, the challenge often lies in the transfer of multi-source heterogeneous data, which

is lacking in research. The difficulty in aligning heterogeneous data at the feature level is a significant issue. The proposed Multi-CDT model in this study successfully circumvents the information loss due to feature modality differences by performing knowledge transfer at the decision level.

Specifically, the Multi-CDT model transfers geometric features from AIS information that are not easily changed due to differences in image imaging principles to the remote sensing image, thereby extracting features with strong discriminative power and significantly improving the robustness and generalization ability of the model. Experimental results show that the Multi-CDT model outperforms traditional models on both the source and target domains. This not only demonstrates the superiority of the model in complex scenarios but also provides new insights for future domain adaptation research. The core of the research lies in finding the commonalities and differences between different data sources, especially when dealing with heterogeneous data, lies in effectively achieving knowledge transfer at the decision level as a topic worth further exploration.

Furthermore, in this study, the classifier trained for AIS information uses XGBoost. In the future, more advanced models can be considered as teacher models. With the rapid development of multi-modal sensor technology, enhancing the fusion and interaction capabilities of multi-source data will become a direction worth exploring in depth. For example, exploring more intelligent multi-modal fusion strategies, such as adaptive weighting or attention mechanisms, to dynamically adjust the contributions of different data sources can further enhance the flexibility and adaptability of the model.

#### 4.4.3. Small Target Vessel Anomaly Analysis and Judgment Task

This paper further discusses and judges the abnormal situation of matching failure and proposes the SVIADF. It analyzes and discusses the abnormal situation of matching failure and takes corresponding measures. We conducted tests using data from Hainan Satellite 1 and Sentinel-1. On the one hand, it shows that this framework can flexibly display the position of ships and provide warnings based on the matching situation of AIS data, which is applicable to both the conventional situation with AIS information and the scenario where AIS information is missing. On the other hand, it verifies the adaptability and robustness of this method under different data sources, which can be widely applied to the fusion and analysis between different remote sensing platforms and AIS data sources.

In the future research, we will focus on building a larger database to further enrich the recognition dataset of optical remote sensing images so as to improve the detection accuracy of the model in different seas, different ship types, and complex environments. This will provide more powerful data support for the identification of small target ships, promote the development of technology in the field of optical remote sensing images, and provide more comprehensive technical support for maritime monitoring and management.

At the same time, in order to ensure real-time requirements, we will continue to optimize the computational efficiency to ensure that the model can still complete inference and analysis in a very short time, even in the case of large-scale data processing. By optimizing the algorithm structure and reasoning process, we believe that this model can adapt to the processing requirements of larger scale data while ensuring high Precision and high Recall so as to improve the accuracy and efficiency of real-time monitoring and anomaly detection, as well as provide more reliable and efficient support for future Marine safety monitoring systems.

In addition, the issue of false positives in busy maritime environments and the inconsistency of AIS data have always been key challenges in maritime monitoring. To further enhance detection capabilities, we plan to expand the depth and breadth of multi-source information fusion, adopt more diverse data sources such as radar and lidar data to reduce

the occurrence of false positives, and enhance the adaptability of the model. At the same time, we will introduce a more refined data repair mechanism to ensure the quality and integrity of AIS data, thereby improving the accuracy and reliability of the entire system.

Finally, the adaptability of the model is also one of the key directions in our future work. We plan to carry out more tests in different geographical areas and environmental conditions to evaluate the performance of the model in variable sea states and further optimize the algorithm based on the test results to enhance its adaptability in different application scenarios. This will help to improve the practical application value of the model in complex and variable environments so that it can deal with more diverse maritime surveillance tasks.

## 5. Conclusions

Based on optical remote sensing images and AIS information, this study constructed a multi-source information fusion framework for small vessel identification and anomaly detection (SVIADF) and realized high-Precision monitoring and analysis of complex maritime scenes through two steps of detection and classification. This framework not only fills the research gap in the identification and abnormal judgment of small target ships under optical remote sensing in academic research but also provides a feasible and operational innovative solution for maritime safety supervision, fishery resource protection, environmental monitoring, and emergency response in the practical application level. In order to improve the practical application value of the framework, we paid special attention to its computational efficiency in large-scale data processing. In the whole multi-source information fusion framework, by optimizing the model structure and reasoning process, it can ensure that the efficient computing power can be maintained when complex image data are analyzed. The final average running time of each image's reasoning prediction, matching, displaying and saving results is 0.083 s, which meets the requirements of real-time monitoring.

In the detection task, the proposed YOLOv8x-CA-CFAR method combines the deep learning detection (YOLOv8x) with the classical CFAR technology to achieve the accurate secondary screening of suspicious targets. The experimental results show that the Recall rate and F1-score of YOLOv8X-CA-CFAR are 2.9%and 1.13% higher than those of YOLOv8x alone, which effectively improves the detection efficiency and accuracy of small target ships. This performance improvement will help strengthen real-time ship monitoring and early risk warning capabilities, as well as provide more reliable data basis for shipping management departments, maritime law enforcement agencies and relevant decision makers, thereby enhancing traffic safety and reducing accident risks.

In the classification task, aiming at the gaps in the research of small target ship classification under optical remote sensing, this study used the unsupervised domain adaptive learning method to successfully transfer the SAR–AIS data features to the optical remote sensing image so as to construct a small target ship classification model suitable for optical remote sensing conditions. Through the proposed multi-source heterogeneous domain adaptation algorithm Multi-CDT, the accuracy and robustness of the model in the complex marine environment are enhanced. Compared to the CDTrans method, Multi-CDT not only retained more classification categories but also improves the accuracy of the target domain by 0.32%, which means that the model has stronger adaptability in the face of diverse ship types and complex sea conditions. This provides an effective technical path for areas lacking SAR data and provides strong support for formulating scientific maritime resource management and control strategies, strengthening maritime safety inspection and management, as well as improving the timeliness and scientification of emergency command and environmental risk assessment.

All in all, the framework and method proposed in this study have not only made solid technical progress at the academic level but also are expected to provide important technical reference and basic support for shipping regulation, resource management, and safety assurance in the South China Sea and other related waters at the application level. This research result will help to promote the further development and practice of marine remote sensing technology in multi-source data fusion and complex environment adaptation, as well as lay a solid foundation for accurate and reliable marine monitoring and decision support in a wider sea area and more severe application conditions in the future.

**Author Contributions:** Conceptualization, L.C. and Z.H.; methodology, L.C.; software, L.C.; validation, J.C. and Y.S.; formal analysis, L.C., Z.H. and J.C.; investigation, J.C. and Y.S.; resources, Z.H.; data curation, L.C. and J.C.; writing—original draft preparation, L.C.; writing—review and editing, Z.H. and Y.S.; visualization, L.C.; supervision, Z.H.; project administration, Z.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the National Natural Science Foundation of China (62361024 and 62161010), the Key Research and Development Project of Hainan Province (ZDYF2022GXJS348 and ZDYF2024GXJS021), and the Hainan University Education and Teaching Reform Research Project (hdjy2421, hdjy2220). The authors would like to thank the referees for their constructive suggestions.

**Data Availability Statement:** The original contributions presented in the study are included in the article, and further inquiries can be directed to the corresponding author.

**Acknowledgments:** We thank the State Key Laboratory of Marine Resources Utilization in the South China Sea and Sanya Research Center, the Institute of Remote Sensing and Earth Sciences, and the Chinese Academy of Sciences for their support. The authors would also like to thank anonymous reviewers for their valuable comments on the manuscript.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Farahnakian, F.; Nicolas, F.; Farahnakian, F.; Nevalainen, P.; Sheikh, J.; Heikkonen, J.; Raduly-Baka, C. A comprehensive study of clustering-based techniques for detecting abnormal vessel behavior. *Remote Sens.* **2023**, *15*, 1477. [CrossRef]
2. Gu, Y.; Hu, Z.; Zhao, Y.; Liao, J.; Zhang, W. MFGTN: A multi-modal fast gated transformer for identifying single trawl marine fishing vessel. *Ocean Eng.* **2024**, *303*, 117711. [CrossRef]
3. Tassetti, A.; Ferrà, C.; Fabi, G. Rating the effectiveness of fishery-regulated areas with AIS data. *Ocean Coast. Manag.* **2019**, *175*, 90–97. [CrossRef]
4. Ferrà, C.; Tassetti, A.N.; Armelloni, E.N.; Galdelli, A.; Scarcella, G.; Fabi, G. Using AIS to attempt a quantitative evaluation of unobserved trawling activity in the Mediterranean Sea. *Front. Mar. Sci.* **2020**, *7*, 580612. [CrossRef]
5. Zhang, T.; Zhao, S.; Cheng, B.; Chen, J. Detection of ais closing behavior and mmsi spoofing behavior of ships based on spatiotemporal data. *Remote Sens.* **2020**, *12*, 702. [CrossRef]
6. Zhang, T.; Liu, C.; Wen, B. Abnormal Ship Behavior Detection after the Closure of AIS Based on Radar Data. *Res. Sq.* **2021**, *3*. . [CrossRef]
7. Zhang, H.; Weijian, L.; Zhang, Q.; Taiyong, F. A robust joint frequency spectrum and power allocation strategy in a coexisting radar and communication system. *Chin. J. Aeronaut.* **2024**, *37*, 393–409. [CrossRef]
8. Zhang, H.; Liu, W.; Zhang, Q.; Liu, B. Joint Customer Assignment, Power Allocation, and Subchannel Allocation in a UAV-Based Joint Radar and Communication Network. *IEEE Internet Things J.* **2024**, *11*, 29643–29660. [CrossRef]
9. Kurekin, A.A.; Loveday, B.R.; Clements, O.; Quartly, G.D.; Miller, P.I.; Wiafe, G.; Adu Agyekum, K. Operational monitoring of illegal fishing in Ghana through exploitation of satellite earth observation and AIS data. *Remote Sens.* **2019**, *11*, 293. [CrossRef]
10. Galdelli, A.; Mancini, A.; Ferrà, C.; Tassetti, A.N. A synergic integration of AIS data and SAR imagery to monitor fisheries and detect suspicious activities. *Sensors* **2021**, *21*, 2756. [CrossRef]
11. Morando, E.; Daffinà, F.C.; Stahl, T.; Corvino, M.M.; Pratola, C. Multi-Sensor Data Fusion to Enhance Maritime Situational Awareness. In Proceedings of the IGARSS 2023–2023 IEEE International Geoscience and Remote Sensing Symposium, Pasadena, CA, USA, 16–21 July 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 6829–6831.
12. Wang, X.; Song, X.; Zhao, Y. Identification and Positioning of Abnormal Maritime Targets Based on AIS and Remote-Sensing Image Fusion. *Sensors* **2024**, *24*, 2443. [CrossRef] [PubMed]

13. Reggiannini, M.; Righi, M.; Tampucci, M.; Lo Duca, A.; Bacciu, C.; Bedini, L.; D'Errico, A.; Di Paola, C.; Marchetti, A.; Martinelli, M.; et al. Remote sensing for maritime prompt monitoring. *J. Mar. Sci. Eng.* **2019**, *7*, 202. [CrossRef]
14. Rodger, M.; Guida, R. Classification-aided SAR and AIS data fusion for space-based maritime surveillance. *Remote Sens.* **2020**, *13*, 104. [CrossRef]
15. Goldstein, J.S.; Reed, I.S. Theory of partially adaptive radar. *IEEE Trans. Aerosp. Electron. Syst.* **1997**, *33*, 1309–1325. [CrossRef]
16. Gandhi, P.P.; Kassam, S.A. Analysis of CFAR processors in nonhomogeneous background. *IEEE Trans. Aerosp. Electron. Syst.* **1988**, *24*, 427–445. [CrossRef]
17. Xu, C.; Li, Y.; Ji, C.; Huang, Y.; Wang, H.; Xia, Y. An improved CFAR algorithm for target detection. In Proceedings of the 2017 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), Xiamen, China, 6–9 November 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 883–888.
18. Rohling, H. Radar CFAR thresholding in clutter and multiple target situations. *IEEE Trans. Aerosp. Electron. Syst.* **1983**, *19*, 608–621. [CrossRef]
19. Fu, K.; Kuang, G.; Yu, W. A Method for Shadow and Target Detection in Synthetic Aperture Radar Images. *J. Softw.* **2002**, 818–826. [CrossRef]
20. Dai, H.; Du, L.; Wang, Y.; Wang, Z. A modified CFAR algorithm based on object proposals for ship target detection in SAR images. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1925–1929. [CrossRef]
21. Kang, M.; Leng, X.; Lin, Z.; Ji, K. A modified faster R-CNN based on CFAR algorithm for SAR ship detection. In Proceedings of the 2017 International Workshop on Remote Sensing with Intelligent Processing (RSIP), Shanghai, China, 18–21 May 2017; IEEE: Piscataway, NJ, USA, 2017, pp. 1–4.
22. Liu, S.Y.; Chen, L.; Hou, X.; Chen, H. Research on Target Detectability Enhancement Technology for SAR Remote Sensing Images. In Proceedings of the 14th National Conference on Signal and Intelligent Information Processing and Applications, Beijing, China, 2021; pp. 301–309.
23. Zhou, W.; Jia, H.; Xiao, X.; Xu, F. Sar Ship Detection Network Incorporating CFAR Preprocessing. In Proceedings of the IGARSS 2022–2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 17–22 July 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 2151–2154.
24. Tang, T. Research on SAR Target Detection Combining Constant False Alarm Rate Detection and Deep Networks. Master's Thesis, Xidian University, Xi'an, China, 2022.
25. Liu, Z.; Yuan, L.; Weng, L.; Yang, Y. A high resolution optical satellite image dataset for ship recognition and some new baselines. In Proceedings of the International Conference on Pattern Recognition Applications and Methods, Porto, Portugal, 24–26 February 2017; SciTePress: Setúbal, Portugal, 2017; Volume 2, pp. 324–331.
26. Xia, G.S.; Bai, X.; Ding, J.; Zhu, Z.; Belongie, S.; Luo, J.; Datcu, M.; Pelillo, M.; Zhang, L. DOTA: A large-scale dataset for object detection in aerial images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3974–3983.
27. Li, K.; Wan, G.; Cheng, G.; Meng, L.; Han, J. Object detection in optical remote sensing images: A survey and a new benchmark. *ISPRS J. Photogramm. Remote Sens.* **2020**, *159*, 296–307. [CrossRef]
28. Cheng, G.; Yuan, X.; Yao, X.; Yan, K.; Zeng, Q.; Xie, X.; Han, J. Towards large-scale small object detection: Survey and benchmarks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 13467–13488. [CrossRef]
29. Chen, J.; Hu, Z.; Wu, W.; Zhao, Y.; Huang, B. LKPF-YOLO: A Small Target Ship Detection Method for Marine Wide-Area Remote Sensing Images. *IEEE Trans. Aerosp. Electron. Syst.* **2024**, 1–15. [CrossRef]
30. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
31. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path aggregation network for instance segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8759–8768.
32. Liu, Z.; Hu, H.; Lin, Y.; Yao, Z.; Xie, Z.; Wei, Y.; Ning, J.; Cao, Y.; Zhang, Z.; Dong, L.; et al. Swin transformer v2: Scaling up capacity and resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 12009–12019.
33. Wang, W.; Xie, E.; Li, X.; Fan, D.P.; Song, K.; Liang, D.; Lu, T.; Luo, P.; Shao, L. Pvt v2: Improved baselines with pyramid vision transformer. *Comput. Vis. Media* **2022**, *8*, 415–424. [CrossRef]
34. Jocher, G.; Chaurasia, A.; Qiu, J.Y. By Ultralytics (2023). 2023. Available online: https://github.com/ultralytics/ultralytics (accessed on 29 November 2023).
35. Huo, B.; Li, C.; Zhang, J.; Xue, Y.; Lin, Z. SAFF-SSD: Self-attention combined feature fusion-based SSD for small object detection in remote sensing. *Remote Sens.* **2023**, *15*, 3027. [CrossRef]
36. Fan, X.; Hu, Z.; Zhao, Y.; Chen, J.; Wei, T.; Huang, Z. A small ship object detection method for satellite remote sensing data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 11886–11898. [CrossRef]

37. Lin, Y.; Li, J.; Wei, S.; Liu, S. OptiShipNet: Efficient Ship Detection in Complex Marine Environments Using Optical Remote Sensing Images. *J. Mar. Sci. Eng.* **2024**, *12*, 1786. [CrossRef]
38. Pelich, R.; Chini, M.; Hostache, R.; Matgen, P.; Lopez-Martinez, C.; Nuevo, M.; Ries, P.; Eiden, G. Large-scale automatic vessel monitoring based on dual-polarization sentinel-1 and AIS data. *Remote Sens.* **2019**, *11*, 1078. [CrossRef]
39. Hou, X.; Ao, W.; Song, Q.; Lai, J.; Wang, H.; Xu, F. FUSAR-Ship: Building a high-resolution SAR-AIS matchup dataset of Gaofen-3 for ship detection and recognition. *Sci. China Inf. Sci.* **2020**, *63*, 1–19. [CrossRef]
40. Yan, Z.; Song, X.; Yang, L.; Wang, Y. Ship classification in synthetic aperture radar images based on multiple classifiers ensemble learning and automatic identification system data transfer learning. *Remote Sens.* **2022**, *14*, 5288. [CrossRef]
41. Long, M.; Cao, Y.; Wang, J.; Jordan, M. Learning transferable features with deep adaptation networks. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 97–105.
42. Sun, B.; Feng, J.; Saenko, K. Return of frustratingly easy domain adaptation. In Proceedings of the AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016; Volume 30.
43. Long, M.; Zhu, H.; Wang, J.; Jordan, M.I. Deep transfer learning with joint adaptation networks. In Proceedings of the International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; pp. 2208–2217.
44. Wu, G. Research on the Application of Transfer Learning in Image Classification. Master's Thesis, Anhui University, Hefei, China, 2017.
45. Fu, J. Research on Deep Transfer Learning Algorithms and Their Applications. Master's Thesis, Nanjing University of Posts and Telecommunications, Nanjing, China, 2020.
46. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *27*.
47. Ganin, Y.; Ustinova, E.; Ajakan, H.; Germain, P.; Larochelle, H.; Laviolette, F.; March, M.; Lempitsky, V. Domain-adversarial training of neural networks. *J. Mach. Learn. Res.* **2016**, *17*, 1–35.
48. Ganin, Y.; Lempitsky, V. Unsupervised domain adaptation by backpropagation. In Proceedings of the International Conference on Machine Learning, Lille France, 6–11 July 2015; pp. 1180–1189.
49. Xu, R.; Li, G.; Yang, J.; Lin, L. Larger norm more transferable: An adaptive feature norm approach for unsupervised domain adaptation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 1426–1435.
50. Xu, C. Research on SAR Image Ship Detection Algorithm Based on Optical Domain Adaptation. Master's Thesis, University of Chinese Academy of Sciences (Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences), Xi'an, China, 2022.
51. Saito, K.; Watanabe, K.; Ushiku, Y.; Harada, T. Maximum classifier discrepancy for unsupervised domain adaptation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3723–3732.
52. Chen, X.; Wang, S.; Long, M.; Wang, J. Transferability vs. discriminability: Batch spectral penalization for adversarial domain adaptation. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 10–15 June 2019; pp. 1081–1090.
53. Xu, T.; Chen, W.; Wang, P.; Wang, F.; Li, H.; Jin, R. Cdtrans: Cross-domain transformer for unsupervised domain adaptation. *arXiv* **2021**, arXiv:2109.06165.
54. Zhang, J. Research on SAR Image Detection and Recognition Methods Integrating Domain Knowledge and Deep Learning. Ph.D. Thesis, Xidian University, Xi'an, China, 2022.
55. Zhao, C. Target Detection and Recognition Technology Using Joint SAR and Optical Images. Master's Thesis, University of Electronic Science and Technology of China, Chengdu, China, 2023.
56. Fei, X.; Guo, M.; Li, Y.; Yu, R.; Sun, L. ACDF-YOLO: Attentive and Cross-Differential Fusion Network for Multimodal Remote Sensing Object Detection. *Remote Sens.* **2024**, *16*, 3532. [CrossRef]
57. Tsai, Y.H.H.; Bai, S.; Liang, P.P.; Kolter, J.Z.; Morency, L.P.; Salakhutdinov, R. Multimodal transformer for unaligned multimodal language sequences. In Proceedings of the Conference. Association for computational linguistics. Meeting, Florence, Italy, 28 July–2 August 2019; Volume 2019, p. 6558.
58. Zhang, D.; Ye, M.; Liu, Y.; Xiong, L.; Zhou, L. Multi-source unsupervised domain adaptation for object detection. *Inf. Fusion* **2022**, *78*, 138–148. [CrossRef]
59. Xie, H. Research on Remote Sensing Image Scene Classification Based on Deep Transfer Learning. Master's Thesis, Harbin Institute of Technology, Harbin, China, 2021.
60. Xian, S.; Zhirui, W.; Yuanrui, S.; Wenhui, D.; Yue, Z.; Kun, F. AIR-SARShip-1.0: High-resolution SAR ship detection dataset. *Radar J.* **2019**, *8*, 852–863.
61. Saidi, S.; Idbraim, S.; Karmoud, Y.; Masse, A.; Arbelo, M. Deep-Learning for Change Detection Using Multi-Modal Fusion of Remote Sensing Images: A Review. *Remote Sens.* **2024**, *16*, 3852. [CrossRef]

62. Liu, F.; Zhang, G.; Lu, J. Multisource heterogeneous unsupervised domain adaptation via fuzzy relation neural networks. *IEEE Trans. Fuzzy Syst.* **2020**, *29*, 3308–3322. [CrossRef]
63. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
64. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1137–1149. [CrossRef]
65. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. SSD: Single shot multibox detector. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Cham, Switzerland, 2016; pp. 21–37.
66. Jocher, G.; Stoken, A.; Chaurasia, A.; Borovec, J.; Kwon, Y.; Michael, K.; Liu, C.; Fang, J.; Skalski, P.; Hogan, A.; et al. Ultralytics/yolov5: v6.0-YOLOv5n 'Nano' models, Roboflow integration, TensorFlow export, OpenCV DNN support. *Zenodo* **2021**. [CrossRef]
67. Wang, C.-Y.; Bochkovskiy, A.; Liao, H.-Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, 20–24 June 2023; pp. 7464–7475.
68. Li, X.; Wang, W.; Wu, L.; Chen, S.; Hu, X.; Li, J.; Tang, J.; Yang, J. Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 21002–21012.
69. Van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
70. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 618–626.