remote sensing

MDPI

Article

Maritime Target Radar Detection and Tracking via DTNet Transfer Learning Using Multi-Frame Images

Xiaoyang He [1], Xiaolong Chen [2,*,†], Xiaolin Du [1,†], Xinghai Wang [2], Shuwen Xu [3] and Jian Guan [2]

[1] School of Computer and Control Engineering, Yantai University, Yantai 264001, China; hexiaoyang@s.ytu.edu.cn (X.H.); duxiaolin168@ytu.edu.cn (X.D.)
[2] Marine Target Detection Research Group, Naval Aviation University, Yantai 264001, China; marshallplan@126.com (X.W.); guanjian_68@163.com (J.G.)
[3] School of Electronic Engineering, Xidian University, Xi'an 710071, China; swxu@mail.xidian.edu.cn
* Correspondence: cxlcxl1209@163.com
† These authors contributed equally to this work.

Abstract: Traditional detection and tracking methods struggle with the complex and dynamic maritime environment due to their poor generalization capabilities. To address this, this paper improves the YOLOv5 network by integrating Transformer and a Convolutional Block Attention Module (CBAM) with the multi-frame image information obtained from radar scans. It proposes a detection and tracking method based on the Detection Tracking Network (DTNet), which leverages transfer learning and the DeepSORT tracking algorithm, enhancing the detection capabilities of the model across various maritime environments. First, radar echoes are preprocessed to create a dataset of Plan Position Indicator (PPI) images for different marine conditions. An integrated network for detecting and tracking maritime targets is then designed, utilizing the feature differences between moving targets and sea clutter, along with the coherence of inter-frame information for moving targets, to achieve multi-target detection and tracking. The proposed method was validated on real maritime targets, achieving a precision of 99.06%, which is a 7.36 percentage point improvement over the original YOLOv5, demonstrating superior detection and tracking performance. Additionally, the impact of maritime regions and weather conditions is discussed, showing that, when transferring from Region I to Regions II and III, the precision reached 92.2% and 89%, respectively, and, when facing rainy weather, although there was interference from the sea clutter and rain clutter, the precision was still able to reach 82.4%, indicating strong generalization capabilities compared to the original YOLOv5 network.

Keywords: maritime moving targets; detection and tracking; radar images; deep learning; inter-frame information

check for updates

Academic Editor: Andrzej Stateczny

Received: 12 December 2024
Revised: 27 January 2025
Accepted: 19 February 2025
Published: 27 February 2025

Citation: He, X.; Chen, X.; Du, X.; Wang, X.; Xu, S.; Guan, J. Maritime Target Radar Detection and Tracking via DTNet Transfer Learning Using Multi-Frame Images. *Remote Sens.* **2025**, *17*, 836. https://doi.org/10.3390/rs17050836

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Radar, as an effective tool for maritime monitoring, is widely applied in various fields such as maritime transport, rescue operations, and fishery detection [1–3]. Among these applications, target detection and tracking are recognized as the primary tasks of maritime surveillance radars. Traditional target detection methods primarily utilize manually designed scale-invariant features [4] and histograms of oriented gradients (HOGs) [5]. However, these manually crafted features often exhibit insufficient robustness against the diverse variations in input data, making them challenging to apply effectively to the complex perception tasks common in maritime environments.

In the field of target tracking, conventional methods often employ correlation filter algorithms, and their improved versions are then used to perform correlation operations on candidate region feature maps. Although these methods enhance tracking efficiency, they typically focus only on the correlation features present between adjacent frames. As a result, when a target undergoes deformation or experiences occlusion, these algorithms are prone to losing track of the target [6–9]. Consequently, traditional methods struggle to achieve high-performance detection and tracking, frequently leading to missed detections or to a loss of targets [10,11]. Additionally, these model-based methods are often limited to specific conditions, which poses challenges in maritime environments that are characterized by significant clutter, the diverse motion characteristics of targets, and unstable tracking scenarios.

As deep learning technologies continue to develop, their applications in the field of radar are also expanding. For example, Chen et al. proposed a controllable false alarm target detection method based on LeNet and feature combination, achieving better detection performance compared to conventional CFAR algorithms [12–16]. Building upon this, Qu et al. introduced the ResNet backbone structure and asymmetric convolution into the design of neural networks, enhancing their detection capabilities [17]. Furthermore, Yan et al. proposed a novel method for target detection in coastal radar systems based on Faster R-CNN, which improves detection precision but faces challenges related to deployment difficulties and slower detection speeds [18].

Typically, target detection algorithms can be classified into two main categories: one-stage and two-stage algorithms. One-stage algorithms, such as You Only Look Once (YOLO) [19] and Single-Shot MultiBox Detector (SSD) [20], directly extract features from images and utilize classifiers or regressors to predict the locations and categories of objects. These algorithms are known for their rapid detection speeds and lower computational complexity, although they may exhibit slightly lower detection precision compared to their two-stage counterparts. In contrast, two-stage algorithms, like Faster R-CNN [21], separate the target detection process into two distinct subtasks: object proposal and object classification. While these two-stage algorithms tend to offer higher precision, they typically operate at slower speeds and entail greater computational complexity.

Moreover, with the ongoing advancements in deep learning, significant progress has been made in deep learning-based target tracking algorithms [22]. For instance, Bertinetto et al. introduced the SiamFC algorithm [23], which is based on a fully convolutional approach to target tracking. This method utilizes a template image of the target and a search region image as inputs, employing a shallow network like AlexNet to extract image features. The feature maps are then convolved with the search region image to generate a response map, ultimately yielding the predicted location of the target. Li et al. introduced the Region Proposal Network (RPN) [24] into the Siamese network tracking framework, proposing the SiamRPN [25] algorithm. This algorithm effectively addresses some of the scale issues inherent in SiamFC, leading to improved tracking precision. However, it also introduces several hyperparameters, which can make the network overly sensitive to these parameters. To overcome the limitations related to spatial invariance, Li et al. proposed the SiamRPN++ [26] tracking algorithm, which advances the feature extraction network from shallow models like AlexNet and GoogleNet to deeper architectures like ResNet [27], fully leveraging the capabilities of deep networks.

Despite these advancements, existing methods in radar image detection face numerous challenges and issues, including complex backgrounds and clutter interference [28]. The multi-frame backgrounds typical of maritime environments and the clutter present in radar echoes can easily result in false detections [29]. Furthermore, the diversity of target motion adds to the complexity as the various forms of movement exhibited by maritime targets

increase the difficulty of tracking [30,31]. While deep learning has improved performance, further enhancements in adaptability are still needed. Additionally, many existing methods fail to incorporate multi-frame information, neglecting the crucial characteristics related to changes in the spatial position of targets over time.

This paper proposes a method for detecting and tracking maritime targets based on inter-frame radar information. It utilizes a combination of the Transformer, Convolutional Block Attention Module (CBAM), and YOLOv5 networks, along with the DeepSORT tracking algorithm and a transfer learning strategy. Compared to existing technologies, the advantages of this method are as follows: it demonstrates effective detection and tracking performance in changing environments and for moving targets, and it focuses on the inter-frame information of moving targets, which includes the trajectory of the target, helping the model better understand dynamic changes. The introduction of the Transformer module allows the model to capture global features and long-range dependencies, which is particularly effective in handling wide-range views. By utilizing the temporal continuity between frames, it reduces the probability of false positives and missed detections, improving detection stability and precision. The CBAM module, a lightweight attention mechanism that is divided into channel and spatial attention, enhances the selectivity of feature representations. It effectively handles background variations, and, by integrating inter-frame information, it can more accurately separate the target from the background. The introduction of transfer learning improves the generalization ability of the network. The unified approach of detection and tracking combines the improved YOLOv5 target detection results with the DeepSORT tracking algorithm, enabling real-time high-precision target detection and tracking. The main objectives of this study are as follows.

1. The concept of inter-frame information is introduced, and real-time data under different sea conditions are collected using a JRC scanning radar. A multi-frame Plan Position Indicator (PPI) image dataset is constructed for various maritime environments.
2. The YOLOv5 network is improved using Transformer and CBAM, and it is combined with the DeepSORT tracking algorithm and a transfer learning strategy to develop the DTNet detection and tracking network.
3. The datasets constructed for different sea conditions are fed into the proposed DTNet network to perform target detection and tracking. By comparing the results with classic networks, the effectiveness of the DTNet network is validated.

This research presents a novel approach for target detection and tracking using radar in maritime environments, significantly enhancing performance. It holds important implications for defense, military operations, and maritime security.

2. Materials and Methods

2.1. Data Acquisition

With its advanced technology, reliable performance, and extensive application range, JRC radars have become a significant choice for users worldwide in marine and terrestrial monitoring [32]. In this study, an X-band incoherent pulse JRC navigation radar, produced by Japan Radio Co., Tokyo, Japan, emits electromagnetic signals. The specific parameters of the radar are shown in Table 1. Figure 1 shows the actual JRC navigation radar, which is used for data collection experiments at the port, coastal boundary, and open sea. It captures the radar echo data of civilian vessels under various sea conditions and weather situations, which are stored as data files in the DAT format. The spatial resolution of radar images is primarily influenced by the radial distance, which plays a critical role in determining the resolution and accuracy of radar images. This is because the time delay and attenuation of signals vary with distance. To achieve optimal results during data collection, the radial

distance is set to approximately 5 nautical miles in coastal boundary and port areas, while, in open-sea regions, it ranges from 10 to 15 nautical miles.

Table 1. JRC navigation radar parameters.

Parameter	Specification
Model	JMA-5312
Transmit Frequency	9410 ± 30 MHz
Speed	24 rpm
Range	0.125–48 nautical miles
Measuring Range	360 degrees in all directions
Transmitting Power	10 kW
Antenna Height	80 m

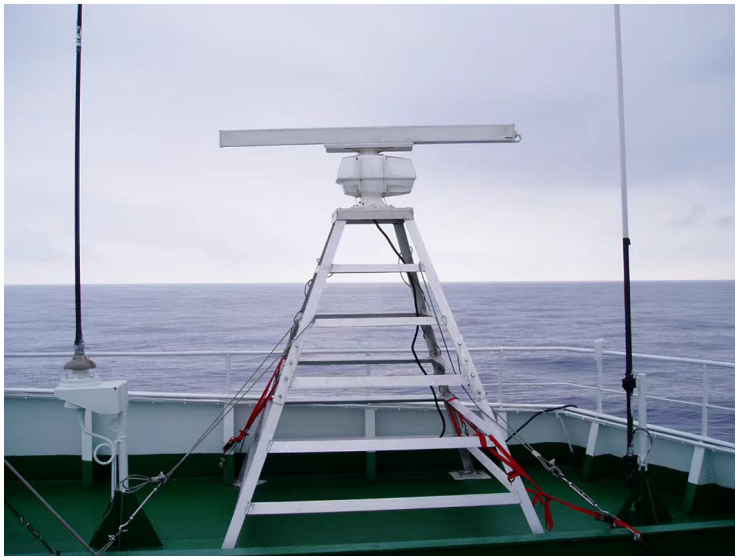


Figure 1. Image of the JRC radar.

2.2. Data Standardization and Preprocessing

By plotting the radar echo data to generate a PPI image, a two-dimensional representation is created that effectively illustrates the distance and direction of the detected targets relative to the geographical position of the radar. The PPI format allows for an intuitive understanding of radar information as it displays targets in a circular layout centered on the radar installation [13]. Each point on the PPI image corresponds to a specific azimuth and range, facilitating easy identification of target locations and their trajectories. The relevant equations are shown below:

$$X(j,:) = \mathrm{D} \times \cos\left(\frac{(90 - \mathrm{A}(j)) \times 2\pi}{360}\right), \tag{1}$$

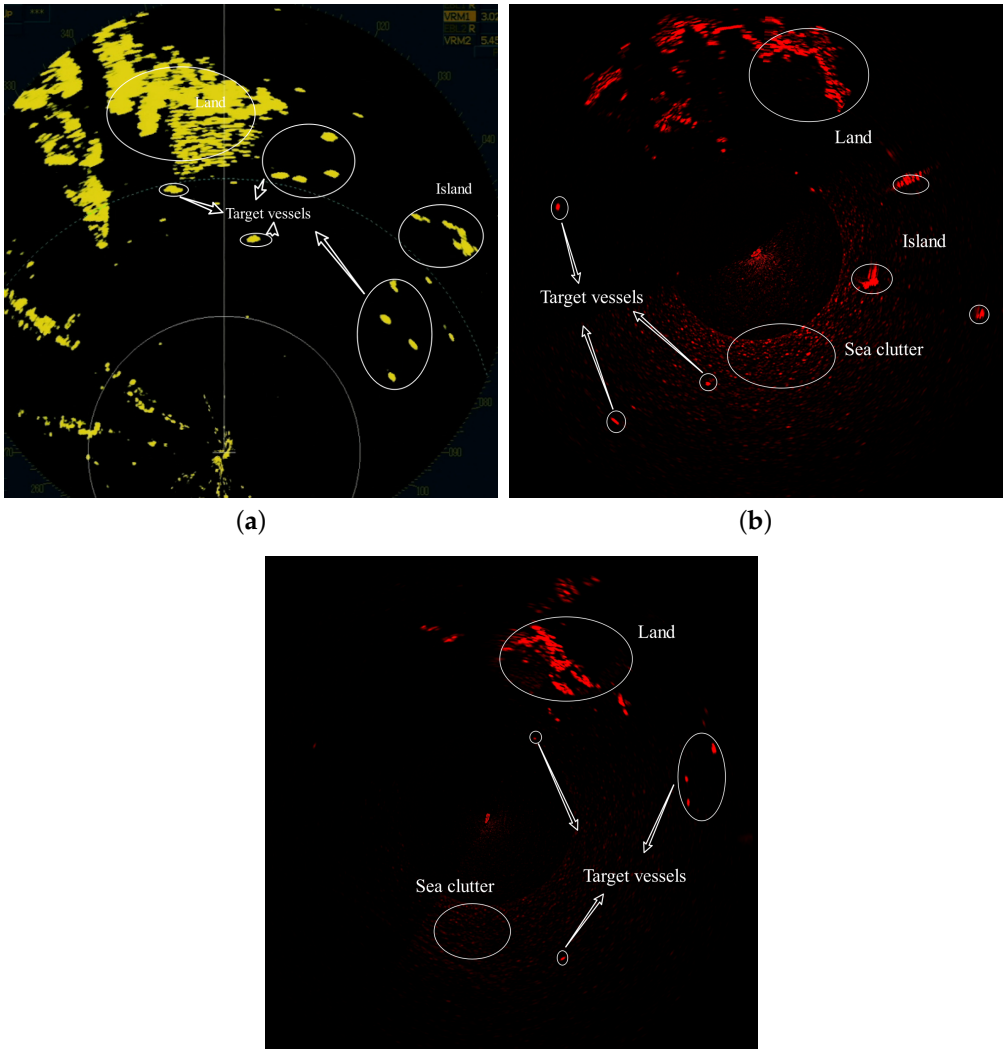
$$Y(j,:) = \mathrm{D} \times \sin\left(\frac{(90 - \mathrm{A}(j)) \times 2\pi}{360}\right). \tag{2}$$

In Equations (1) and (2), D represents the distance grid points, while A denotes the angle data. The variables X and Y refer to the coordinates used when plotting the PPI image. The radar sea surface echo data are collected and preprocessed to construct a navigation radar image database, as shown in Table 2. This database contains PPI images of sea surface echoes under three different maritime environments. The data in Figure 2 correspond to three different maritime areas: the dataset for Maritime Area I includes four types of weather data, consisting of 6591 images. Among them, there are 1762 images for sunny

weather, 1390 images for rainy weather, 1570 images for foggy weather, and 1869 images for windy weather. Sunny weather corresponds to Sea State Levels 1–2, with calm seas and minimal clutter; rainy weather corresponds to Sea State Levels 3–4, with light-to-moderate waves and rain-induced interference; foggy weather corresponds to Sea State Levels 1–2, with calm seas but low visibility and limited clutter; and windy weather corresponds to Sea State Levels 5–6, with significant waves and increased sea clutter. In high sea states (rainy and windy weather), clutter suppression techniques can be applied first to reduce interference, followed by detection and tracking to improve accuracy. Maritime moving targets in the dataset exhibit various motion types, including straight-line motion, turning, maneuvering, and crossing. These targets often undergo shape deformation due to factors such as changes in posture, sea wave interference, and radar resolution limitations. Such deformations adversely affect detection and tracking performance by increasing the difficulty of feature extraction, leading to false detections, missed detections, or trajectory drift. Additionally, deformation complicates the estimation of speed and direction. The training set is predominantly composed of data from Maritime Area I, while the test set includes data from Maritime Areas II and III. Variations in sea conditions and target motion patterns across these areas are the main factors influencing detection and tracking accuracy. Figure 3 corresponds to maritime scenarios under four different weather conditions. The data were collected at a port that is frequently passed by numerous ships. The datasets for Maritime Areas II and III were collected along coastal boundaries and offshore locations, with 59 and 61 images, respectively. The targets collected in this study include transport ships, fishing vessels, etc., as shown in Figure 4. Figure 5 shows the data collected under sunny, rainy, foggy, and windy conditions. It can be observed that the data collected in foggy conditions are similar to that collected on sunny days, indicating that fog has minimal impact on the transmission and reception of radar signals. However, rain and windy conditions show noticeable interference from clutter and sea clutter, which is similar to the target signal, making it prone to misjudgment. In Maritime Area I, a total of 56 targets were detected under four weather conditions. Maritime Area II had 7 targets, and Maritime Area III had 6 targets. Additionally, the training and testing datasets in this study were divided in a 7:3 ratio. In this study, rainfall on rainy days exceeds 10 mm per hour but is less than 50 mm per hour. At sea, this forms areas of dense precipitation, generating rain clutter. Rain clutter is caused by radar signals scattering off raindrops, with the strong echoes from the raindrops potentially masking weaker target echoes, such as those from small boats or floating debris. The intensity of rain clutter is affected by factors such as rainfall rate, raindrop size distribution, and radar frequency. High-frequency radar is particularly susceptible to attenuation and scattering effects from raindrops. Therefore, heavy rain significantly reduces radar signal clarity, decreases target detectability, and increases the likelihood of false alarms, posing a major challenge to maritime radar systems.

Table 2. Dataset composition description.

Data Type	Collection Location	Number of Targets (Count)	Training Set (Images)	Test Set (Images)
Maritime Area I	Port	56	4614	1977
Maritime Area II	Coastal Intersection	7	-	59
Maritime Area III	Offshore	6	-	61



(**a**) (**b**)

(**c**)

Figure 2. Data from different maritime areas. (**a**) Maritime Area I. (**b**) Maritime Area II. (**c**) Maritime Area III.



(**a**) (**b**)

Figure 3. *Cont.*



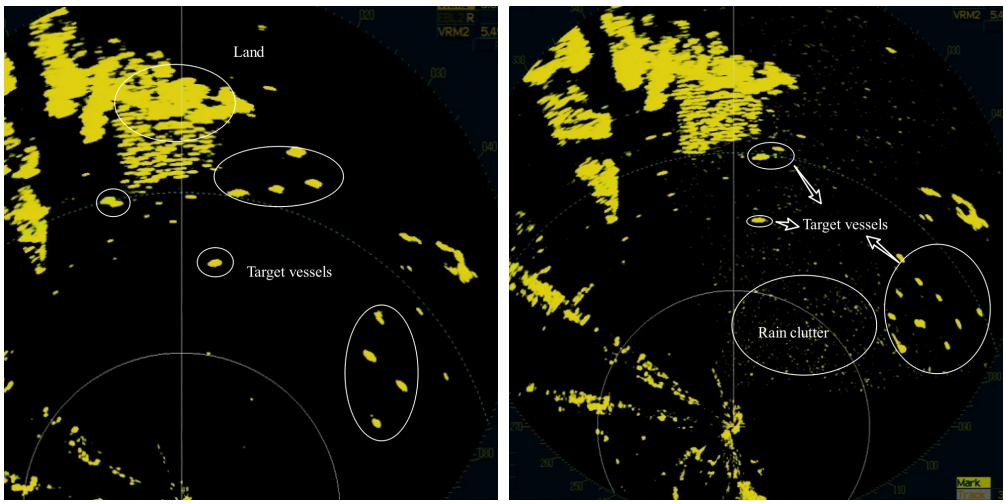
(**c**) (**d**)

Figure 3. The actual sea surface conditions under different weather scenarios. (**a**) Sunny day. (**b**) Rainy day. (**c**) Foggy day. (**d**) Windy day.



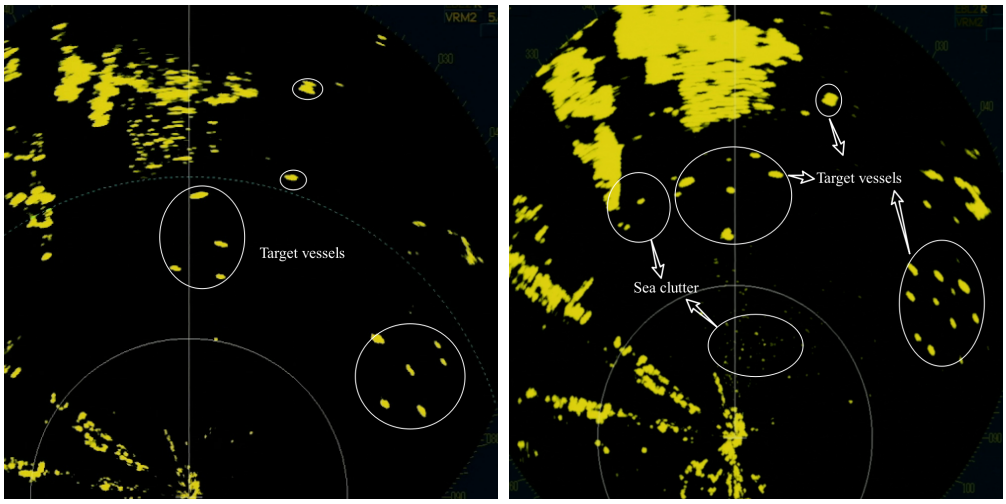
(**a**) (**b**)

Figure 4. Actual images of the target vessels. (**a**) Fishing vessels. (**b**) Transport ship.



(**a**) (**b**)

Figure 5. *Cont.*



(**c**) (**d**)

Figure 5. Different weather data. (**a**) Sunny day. (**b**) Rainy day. (**c**) Foggy day. (**d**) Windy day.

2.3. Model Construction

The main workflow of the proposed DTNet network is illustrated in Figure 6. The specific steps are as follows: First, radar echo data undergo preprocessing to establish PPI images under various marine environments, which serve as the dataset. Next, the model is trained and tested using an inter-frame sequence dataset, resulting in optimal network model parameters. Through a transfer learning strategy, the model parameters that have been trained in the source domain are transferred to different marine data.

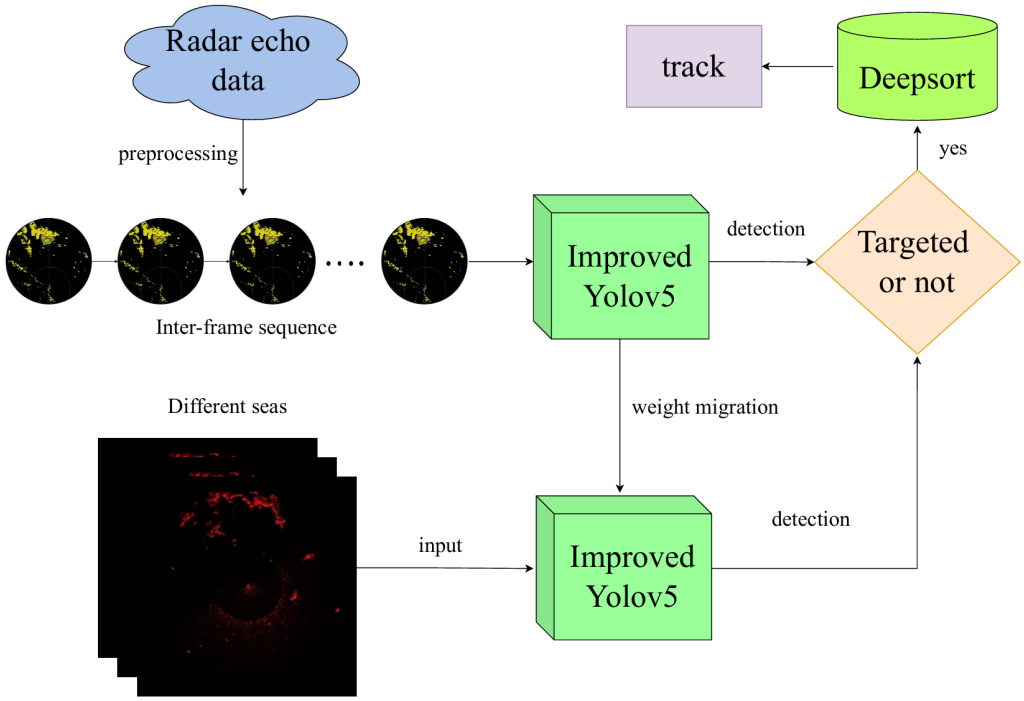


Figure 6. The overall flow of the DTNet network.

The process can be divided into several key stages: preprocessing the radar echo signals through a software processing platform to construct an inter-frame echo image dataset, followed by the development of an improved YOLOv5 detection and tracking network model. This model comprises three modules, as detailed in the subsequent sections: the detection module, the tracking module, and the transfer learning module.

The improved YOLOv5 module integrates Transformer and CBAM with YOLOv5 to process the inter-frame sequence images. The DeepSORT tracking module is employed to achieve radar target tracking. The training set is then inputted to iteratively optimize and train the improved YOLOv5 network.

Subsequently, motion parameter estimation is conducted. The model parameters are transferred and trained, with the pre-trained model parameter files initializing the weights of the network, thereby replacing the original random initialization operation. This process involves global fine tuning, after which the source domain and target domain are modified to enhance the generalization capability of the network.

Through these steps, the DTNet network effectively leverages the characteristics of radar echo data, ensuring accurate detection and tracking of targets in various marine conditions. This approach not only improves performance in specific environments, but also enhances the adaptability of the model to new, unseen data, ultimately contributing to more robust and reliable radar signal processing.

2.3.1. Composition of the Detection Module

The YOLOv5 network stands out for its excellent real-time performance and a well-balanced tradeoff between precision and speed. Compared to other one-stage detection algorithms, YOLOv5 achieves higher detection accuracy in complex scenarios while maintaining a high frame rate, making it ideal for applications with strict real-time requirements. In contrast to two-stage algorithms, YOLOv5 features lower computational overhead and a more lightweight model, making it easier to deploy on resource-constrained devices. To address the challenges of complex environmental backgrounds, clutter or noise interference, extreme target motion, and inter-frame information issues in radar signal detection, this study presents several improvements based on YOLOv5, as illustrated in Figure 7. These enhancements incorporate Transformer and CBAM to boost the performance of the model. The specific improvements are as follows.

(1) Composition of the Transformer.

In radar signal detection, the complex ocean background and clutter interference make feature extraction more challenging, thereby hindering target recognition. To effectively extract the inter-frame information from radar data, the Transformer structure is integrated into the YOLOv5 network. By leveraging the self-attention mechanism in the Transformer, as shown in Figure 8, each frame's feature information is computed, including the query, key, and value. These computations determine the similarity between each frame's features and those of other frames, where the focus is on the frame information most relevant to the current frame. Through the self-attention mechanism, the Transformer facilitates the transfer of information between frames, ensuring that target detection in the current frame is not only solely reliant on its features, but that it also incorporates information from preceding and succeeding frames for more accurate judgments. The self-attention mechanism is defined as follows:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \tag{3}$$

where Q, K, and V are the query, key, and value matrices, respectively, and d_k are the dimensions of the key matrix [33].

In recent years, significant progress has been made in object detection tasks using deep convolutional neural networks [19,20,24,34,35]. By integrating CNN with the Transformer, a convolutional Transformer module is designed, as illustrated in Figure 8 [36]. In this Transformer module, to enhance the feature extraction of the model and capture the motion of targets across multiple frames, a convolutional construction approach is employed. This

enables the network to possess both the local capturing capabilities of CNNs and the global modeling abilities of Transformers. The module is divided into two parts: local modeling and global modeling. In the local modeling phase, to better address issues related to scale variations of targets during tracking, thereby improving algorithm robustness, a deformable convolution layer is utilized to encode local spatial information. This is followed by a convolutional layer that adjusts the number of feature channels. In the global modeling phase, features from the local modeling segment are transformed into non-overlapping image patches. By employing a Transformer to encode the relationships between these patches, the model performs weighted fusion of features at different locations, resulting in richer and more accurate feature representations. This process significantly enhances the precision and robustness of target tracking. Subsequently, the image patches are folded back into feature maps to realize global feature modeling, with a convolutional layer adjusting the feature channel count back to the original size. Furthermore, through a shortcut branch, the features from local modeling are concatenated with those from global modeling. A convolutional layer is then used to fuse these two types of feature information. Finally, the output from the convolutional layer is combined with the input features, resulting in the final output. This residual connection approach optimizes the network architecture and mitigates issues related to information loss and feature degradation. By implementing this convolutional Transformer module, the model achieves improved performance in detecting and tracking targets in dynamic and complex radar environments, thus enhancing the overall effectiveness of the radar signal processing framework.

By leveraging the strengths of Transformer architecture, the model can better focus on relevant features while ignoring irrelevant background noise, thus improving detection precision and robustness. This enhancement is particularly critical in scenarios where targets exhibit rapid movements or are obscured by environmental clutter. The improved model not only facilitates better feature representation, but also allows for a more comprehensive understanding of the temporal dynamics involved in radar signal detection.

(2) YOLOv5 enhancements for ultra-small target detection

Building upon the original YOLOv5 framework, the detector was adjusted to incorporate four distinct detectors, including an additional one specifically designed for the detection of ultra-small targets, as illustrated in Figure 7. In maritime target detection and tracking tasks, targets of varying sizes exhibit different feature representations. This is particularly true for ultra-small targets, which often occupy a minimal number of pixels in the image, resulting in less prominent features that can easily be overlooked.

The improved YOLOv5 enhances the detection structure by adding a dedicated detector for ultra-small targets, allowing for more effective detection of these challenging objects. By focusing on the unique characteristics and spatial constraints associated with ultra-small targets, the modified YOLOv5 can significantly improve detection precision. This targeted approach not only enhances the overall performance of the model in maritime environments, but also addresses the critical challenge of detecting small-scale targets that are often essential for comprehensive situational awareness in various applications.

The introduction of a specialized detector for ultra-small targets enables the model to better adapt to the specific nuances of radar imagery, ensuring that even the most diminutive and often neglected targets are accurately identified and tracked. This advancement contributes to the robustness of the detection system, ultimately facilitating more reliable maritime operations and improving safety in navigational contexts.

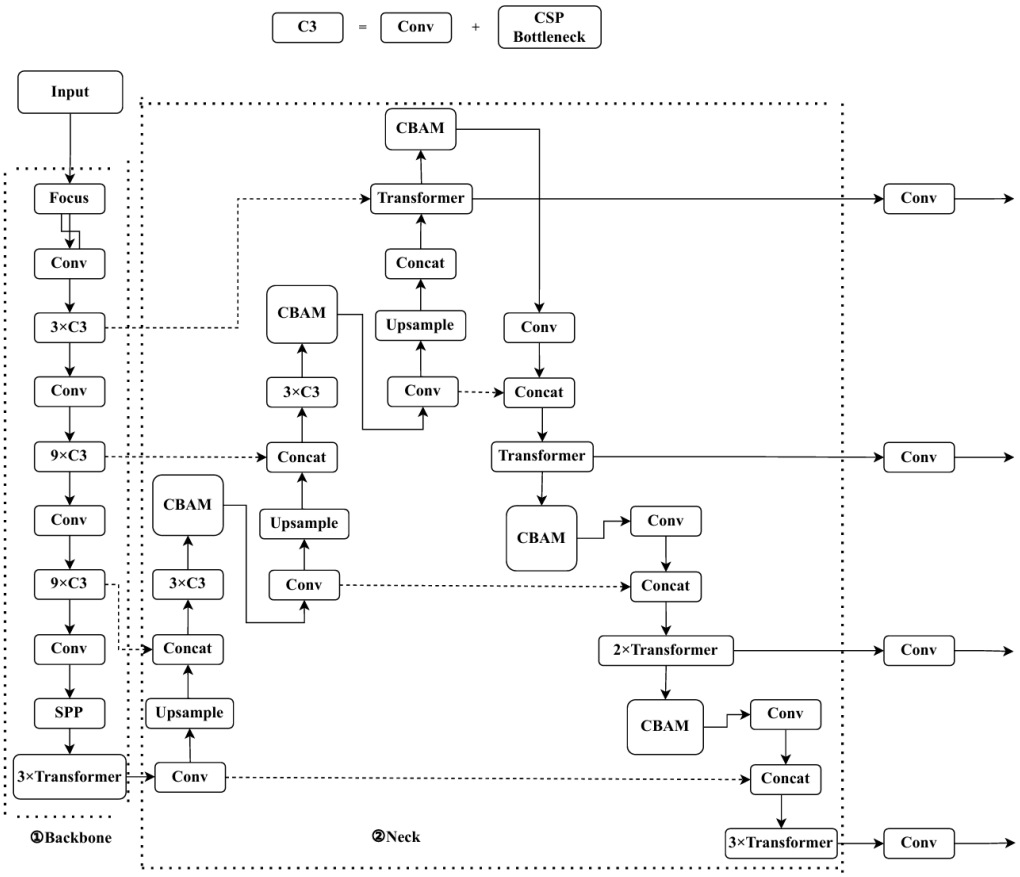


Figure 7. Improved YOLOv5 network.

(3) Composition of the CBAM

In radar signals, extreme target motion and insufficient precision in inter-frame information processing can adversely affect detection performance with respect to the detection of moving targets and complex background handling. By integrating the CBAM module into YOLOv5, attention mechanisms are utilized to enhance the focus of the model on moving targets. The CBAM module utilizes both channel and spatial attention mechanisms, which allows for improved handling of inter-frame information and moving targets within complex backgrounds, ultimately increasing detection precision. Additionally, CBAM effectively alleviates the challenges posed by moving target deformation by combining channel attention and spatial attention mechanisms. Channel attention focuses on the importance of different channels, enhancing those related to target feature variations while suppressing irrelevant features, thereby capturing critical features even when the target undergoes deformation. Spatial attention generates a spatial attention map to dynamically adjust the network's focus on target regions, enabling it to concentrate on significant areas even when the target's shape or outline changes. Moreover, the CBAM module integrates global and local information with dynamically generated attention mechanisms to adjust feature mappings in real time, enhancing the network's robustness to deformed targets. This ensures efficient feature extraction in complex and dynamic scenarios.

These enhancements collectively elevate the performance of the model in detecting radar signals amid composite interference, making it more suitable for practical applications involving complex and high-density target detection scenarios. The neck of the algorithm employs a mixed-domain CBAM. This simple yet effective attention module consists of two main components, as illustrated in Figure 9: channel attention and spatial attention.

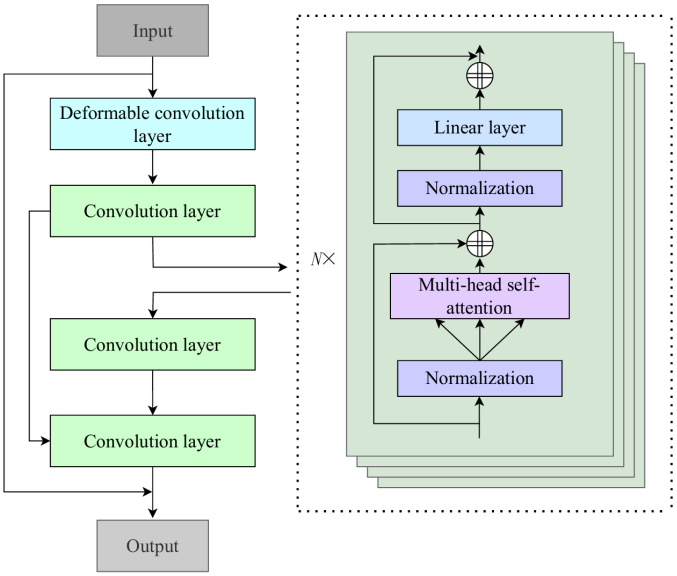


Figure 8. Transformer module.

CBAM operates by inferring the attention for the image across both the channel and spatial dimensions based on intermediate feature maps. Subsequently, the attention layers are combined with the feature maps to perform adaptive feature refinement. The relevant formulas for these operations are as follows:

$$\begin{aligned} M_c(F) &= \sigma(\mathrm{MLP}(\mathrm{AvgPool}(F)) + \mathrm{MLP}(\mathrm{MaxPool}(F))) \\ &= \sigma\left(W_1\left(W_0\left(F^c_{\mathrm{avg}}\right)\right) + W_1(W_0(F^c_{\mathrm{max}}))\right), \end{aligned} \tag{4}$$

$$\begin{aligned} M_s(F) &= \sigma\left(f^{7\times7}([\mathrm{AvgPool}(F); \mathrm{MaxPool}(F)])\right) \\ &= \sigma\left(f^{7\times7}\left(\left[F^s_{\mathrm{avg}}; F^s_{\mathrm{max}}\right]\right)\right). \end{aligned} \tag{5}$$

In the equations, σ represents the sigmoid operation; 7×7 denotes the size of the convolution kernel; F indicates the input feature map; and AvgPool, MaxPool, and MLP correspond to average pooling, max pooling, and multi-layer perceptron, respectively. The CBAM is a lightweight module that can be seamlessly integrated into CNN architectures and trained in an end-to-end manner.

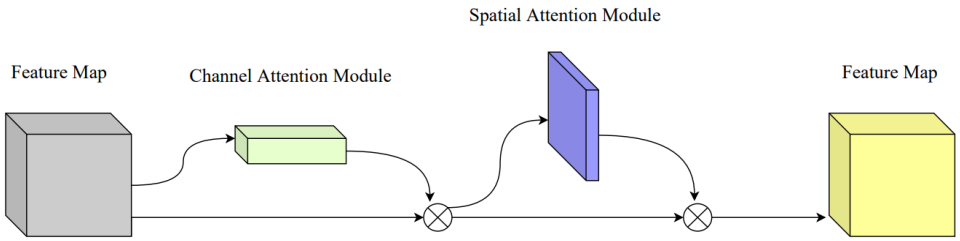


Figure 9. Convolutional block attention module.

2.3.2. Composition of the Tracking Module

In traditional object tracking algorithms, Simple Online and Realtime Tracking (SORT) is favored for its speed, primarily due to its reliance on the Kalman filter for motion estimation and the Hungarian algorithm for data association. However, while SORT excels in terms of speed, its performance is often limited when handling challenges such as target occlusion, appearance variation, and the multi-object tracking of similar targets. This limitation becomes particularly evident in complex scenes, where the SORT algorithm struggles to effectively differentiate and track multiple targets.

To enhance the target tracking performance of the DTNet network in complex environments, this paper integrated the DeepSORT tracking algorithm into the DTNet network. Compared to the original DeepSORT, which mainly relies on the Kalman filter and the Hungarian algorithm to combine target motion information (such as speed and position) with appearance features for data association, this study further optimized the data association strategy, particularly in multi-target scenarios. By enhancing the use of appearance features and introducing more flexible data association algorithms, the model can more effectively handle target similarities, reduce mismatches, and improve performance, especially in complex situations such as target overlap, occlusion, and background interference. The improved YOLOv5 provides static detection information of the targets, while DeepSORT serves as the key module for tracking. By associating and tracking the detection results, it transforms them into dynamic information, forming complete target trajectories. The combination of the two creates a comprehensive detection-tracking system. DeepSORT leverages the powerful feature extraction capabilities of deep learning while maintaining the real-time efficiency of the SORT algorithm. By incorporating a deep feature extraction module and multi-feature fusion techniques, DeepSORT significantly improves the ability to describe targets, particularly in scenarios involving occlusion and appearance changes, as illustrated in Figure 10. The workflow of the DeepSORT algorithm is as follows.

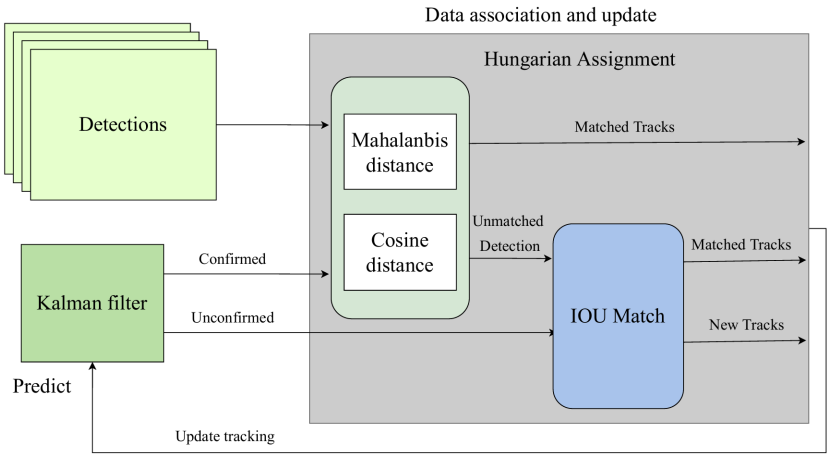


Figure 10. Flowchart of the DeepSORT tracking algorithm.

Initialization: In the first frame, detected targets are identified, and corresponding trajectories are created for each target. The motion variables of the Kalman filter are initialized, allowing for the prediction of corresponding bounding boxes. At this stage, the trajectory state is classified as unconfirmed.

IOU Matching: The bounding boxes from the target detections of the current frame are matched one-to-one with the predicted bounding boxes from the previous frame through Intersection over Union (IOU) calculations, generating a cost matrix for this association. A is the predicted bounding box, and B is the ground truth bounding box. The IoU is calculated as follows:

$$\text{IoU} = \frac{A \cap B}{A \cup B}. \tag{6}$$

Hungarian Algorithm: The cost matrix is used as the input for the Hungarian algorithm to obtain a linear matching result. Given an $n \times n$ matrix C, the objective is to find a permutation σ (i.e., a specific arrangement) that minimizes the total cost:

$$\min \sum_{i-1}^{n} C_{i,\sigma(i)}. \tag{7}$$

Iterative Updates: The steps of IOU matching and the Hungarian algorithm are repeated until a confirmed trajectory is established or the video frames have ended.

Cascaded Matching: For confirmed trajectories, Kalman filter predictions are generated, and the predicted bounding boxes are matched with detections using a cascade approach that incorporates appearance features and motion information.

Re-matching: After the cascaded matching, unmatched trajectories and detections undergo IOU matching again, followed by linear matching using the Hungarian algorithm. The matching results fall into three scenarios that are similar to previous cases.

Cyclic Processing: The steps of cascaded matching and re-matching are repeated until all video frames have been processed.

2.3.3. Transfer Learning Module

First, the parameters of the pre-trained model are used to initialize DTNet with a multi-frame dataset from the source domain, replacing random initialization to improve the starting point for model training, as shown in Figure 11. Next, global fine tuning is performed on the target domain to adapt the model to new task requirements. The process of transfer learning involves migrating knowledge between the source and target domains to enhance the generalization ability of the model. This method not only improves detection precision and robustness, but also accelerates the training process, addresses issues of data scarcity, and effectively manages the complexities of maritime environments, ensuring effective real-time target tracking. The specific steps are as follows: during model transfer, the pre-trained model parameter file initializes the model network weights, replacing the original random initialization, followed by global fine tuning, while the remaining training process continues as usual. The transfer learning studied in this paper focuses on a source domain D_s and a target domain D_t, where $D_s = \{x_i, y_i\}_i^N$ represents the source domain; x_i and y_i represent the data samples and the corresponding moving target category labels, respectively; and the target domain is represented by $D_t = \{x_i, y_i\}_i^{N_t}$. Given the source domain D_s and learning task T_s, as well as the target domain D_t and learning task T_t, the purpose of transfer learning is to acquire knowledge from the source domain D_s and learning task T_s to help improve the learning of the prediction function $f_t(\cdot)$ in the target domain, where $D_s \neq D_t$ or $T_s \neq T_t$. The overall process of model transfer is as follows (where W_s represents the pre-trained weights of the model in the source domain, and W_t represents the training weights of the model in the target domain): $W_s \xrightarrow{s \to t} W_t$.

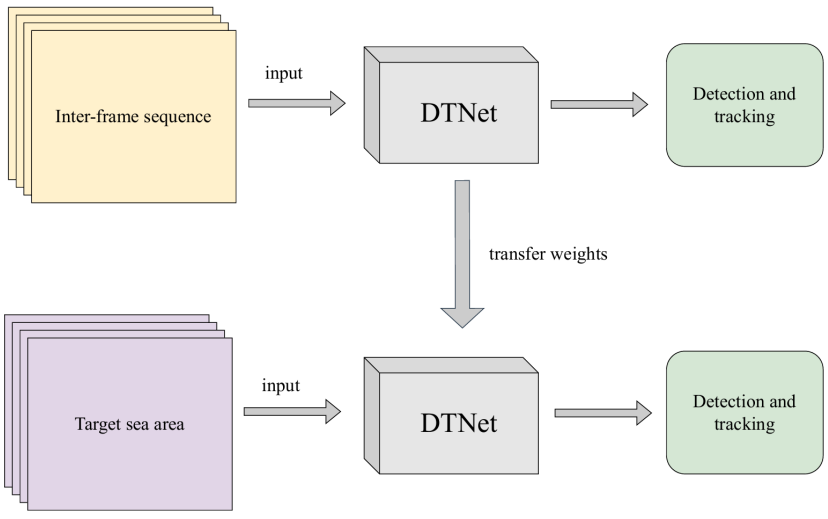


Figure 11. The DTNet transfer learning process.

2.3.4. Target Motion Parameter Estimation

In the DeepSORT algorithm [37], by tracking a moving target over multiple frames, the motion parameters of the target, including speed and heading, can be estimated. The following are the steps and related equations for estimating the target motion parameters using the DeepSORT algorithm. The DeepSORT algorithm uses a Kalman filter to estimate the state of each target. The algorithm assumes that the state vector of the target is as follows:

$$X_k = \left[x_k, y_k, s_k, r_k, v_x, v_y\right]^T, \tag{8}$$

where (x_k, y_k) is the position of the target in the first frame (center point coordinates); s_k denotes the scale (area) of the target; r_k is the aspect ratio of the target; and (v_x, v_y) represents the velocity component of the target.

The Kalman filter prediction can be expressed as follows:

$$x_{k|k-1} = Fx_{k|k-1}, \tag{9}$$

$$P_{k|k-1} = FP_{k|k-1}F^{\mathrm{T}} + Q. \tag{10}$$

Here, $x_{k|k-1}$ represents the predicted state vector (the state prediction at the current time step k), F is the state transition matrix, $P_{k|k-1}$ denotes the predicted error covariance matrix, and Q represents the process noise covariance matrix. The Kalman filter update can be expressed as follows:

$$K_k = P_{k|k-1}H^{\mathrm{T}}\left(HP_{k|k-1} + R\right)^{-1}, \tag{11}$$

$$x_{k|k} = x_{k|k-1} + K_k\left(z_k - Hx_{k|k-1}\right), \tag{12}$$

$$P_{k|k} = (I - K_kH)P_{k|k-1}, \tag{13}$$

where H is the observation matrix; R is the observation noise covariance matrix; K_k is the Kalman gain; I represents the identity matrix, with dimensions that match the matrix K_kH, thus ensuring consistency in matrix operations; z_k is the current observation vector; and $Hx_{k|k-1}$ is the predicted value from the observation model.

With the update step of the Kalman filter, the estimated values of the target velocity components (v_x, v_y) can be obtained:

$$\hat{V}_k = \left[\hat{v}_x, \hat{v}_y\right]^{\mathrm{T}}. \tag{14}$$

Heading θ_k is the direction angle of the target motion, which can be calculated from the following velocity component:

$$\theta_k = \arctan 2\left(\hat{v}_x, \hat{v}_y\right), \tag{15}$$

where arctan 2 is the binary arctangent function for calculating the four-quadrant arctangent angle.

3. Results

3.1. Training Methods

In this paper, the environment architecture used includes Python 3.9, PyCharm 2021, CUDA 11.3, cuDNN 8.2, and PyTorch 1.12, with image processing through the graphical interface of NVIDIA DIGITS. The computer configuration consists of dual E5 processors, an NVIDIA GeForce RTX 3080 GPU, and 32 GB of memory. Based on the convergence of the loss value during training, the training parameters of the model were set as follows:

200 iterations, stochastic gradient descent (SGD) for parameter optimization, a step-down decay strategy, an initial learning rate of 0.001, and a decay rate of 0.1.

(1) Inputting the training dataset to iteratively optimize and train the improved YOLOv5 network.

During the iterative optimization training, the Adam gradient descent method is used via utilizing the parameters θ, gradient g_t, momentum parameter β_1, learning rate parameter β_2, and a constant ϵ. First, an exponential weighted moving average is performed to estimate the first moment m_t and the second moment v_t of the gradient:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t, \tag{16}$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2. \tag{17}$$

Then, the parameters are updated based on the corrected first moment and second moment:

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}, \tag{18}$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t}, \tag{19}$$

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t} + \ell} \hat{m}_t, \tag{20}$$

where t represents the current iteration count and α denotes the learning rate.

(2) Model parameter transfer training

The Adam gradient descent method is used for iterative optimization training. In the task of detecting and tracking targets at sea, transfer learning can enhance model performance in specific environments (such as different maritime areas and weather conditions) through parameter transfer. During the transfer learning process, model pairs are trained, after which the parameters are frozen, and the gradient descent method is applied to update the parameters:

$$\theta_{t+1} = \theta_t - \eta \nabla_\theta L_{\text{DTNet}}, \tag{21}$$

where θ_t is the model parameters at t iterations and η is the learning rate.

When updating frozen layers, if a layer is frozen, its parameters are not updated. Regularization is used to prevent overfitting when the data volume is limited. The relevant formula is as follows:

$$L_{\text{total}} = L_{\text{DTNet}} + \lambda \|\theta\|_2^2, \tag{22}$$

where $\|\theta\|_2^2$ indicates the regularization term, and λ denotes the regularization factor.

3.2. Model Detection and Tracking

This section evaluates the DTNet algorithm model on the Maritime Area I test set for detection and tracking. The proposed model demonstrates low loss values and high precision on the data, successfully performing maritime target detection and tracking tasks on both the training and validation sets. The experimental simulation results, as shown in Figure 12a, present the training loss values. The total loss consists of classification loss (class_loss), regression loss (box_loss), and object confidence loss (obj_loss). It can be observed that, as the number of iterations increases, the training loss stabilizes around 0.02.

Combining deep learning target detection and radar target detection, the evaluation of detection results is conducted using four metrics: recall rate, precision, false alarm rate, and detection rate (FPS). The results are shown in the table. Here, FPS represents the number of

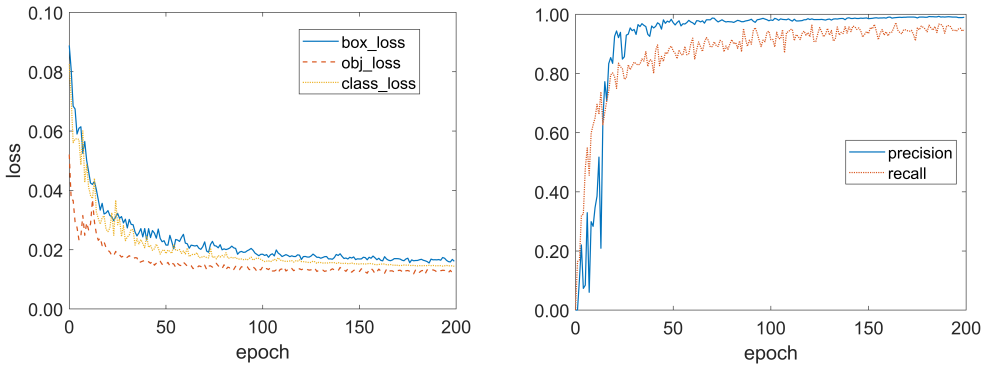
images (frames) recognized within one second. The formulas for recall rate, precision, and false alarm rate are as follows:

$$R = \frac{TP}{TP + FN}, \tag{23}$$

$$P = \frac{TP}{TP + FP}, \tag{24}$$

$$FA = 1 - P, \tag{25}$$

where R represents Recall, P represents Precision, FA stands for false alarm rate, FP (False Positive) indicates the prediction of negative samples as positive, TP (True Positive) indicates the prediction of positive samples as positive, and FN (False Negative) indicates the prediction of positive samples as negative. After training, the dataset from Maritime Area I was used for detection and tracking tests, the results of which are shown in Figure 12b. It can be observed that, as the number of iterations increased, both precision and recall continually rose, ultimately stabilizing at a recall rate of 94.6% and a precision rate of 99.06%, as detailed in Table 3.



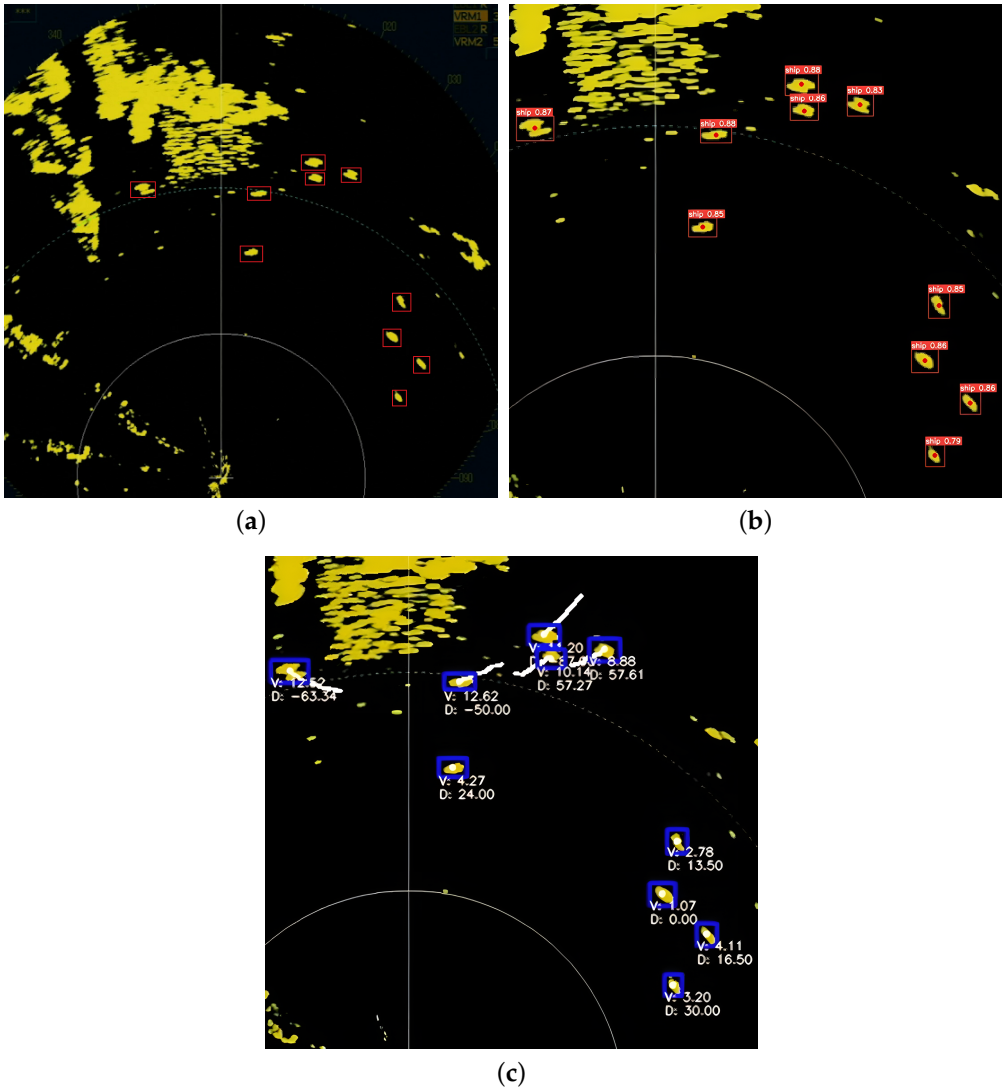
(**a**) (**b**)

Figure 12. Relevant parameter curve. (**a**) Loss function curve. (**b**) Precision–recall curve.

Table 3. Evaluation results.

Algorithm	Recall (%)	Precision (%)	FA (%)	FPS
DTNet	94.6	99.06	0.94	7.64

When testing the detection performance, a random selection from the Maritime Area I test set was input into the network for evaluation. Figure 13a corresponds to the respective position of the target Automatic Identification System (AIS) in Dataset I. Additionally, it should be noted that AIS target information was used as the ground truth in this paper, as well as in the subsequent experiments, in order to derive the detection and tracking probability predicted by the algorithm. The detection results of the network model are shown in Figure 13b, indicating a detection probability of around 80%, which is quite good. For the tracking performance test, the data from Maritime Area I was organized into a video, and the target tracking was accumulated frame by frame to form tracking trajectories, as illustrated in Figure 13c. Here, V represents the relative speed of the target between two frames, and D indicates the motion direction between the two frames.



(**a**) (**b**)

(**c**)

Figure 13. AIS and detection tracking results for Maritime Area I. (**a**) The AIS display of the targets in Maritime Area I. (**b**) Detection performance. (**c**) Tracking performance.

3.3. Comparative Analysis

To validate the performance of the proposed model, comparisons were made with classic algorithms such as Faster R-CNN, RetinaNet, and YOLOv5. In the experiments, the training set of the created maritime target dataset was used as input data for training, and the test set was used for testing and validation. The test results are shown in Figure 14, and the comparison of evaluation metrics is presented in Table 4, with detection results depicted in Figure 15. Additionally, when comparing with AIS, as shown in Figure 13a, it can be observed that there were still detection errors in the results, as shown in Figure 15b–d. On the one hand, some maritime targets have similar features to the background (such as sea clutter or island contours), leading to confusion in the detection algorithm. This issue is particularly prominent in multi-target scenarios, where the similarity between targets and the background or other targets increases the likelihood of false detections or missed detections. On the other hand, in complex maritime scenes, the algorithm may struggle to adapt to certain specific environments due to limitations in its structure, feature extraction capabilities, or loss function design, leading to incorrect detections.

The experimental results show that the DTNet network outperforms classical algorithms in maritime environment tests. Firstly, Faster R-CNN and RetinaNet, which adopt feature pyramid structures (FPN), exhibit certain limitations when dealing with multi-scale

targets with large-scale variations, especially when higher-level features fail to effectively express the characteristics of small targets. Faster R-CNN typically achieves good accuracy, but its performance drops significantly in detecting multi-scale targets, particularly small ones. Although RetinaNet uses focal loss to address the class imbalance issue, it still struggles with detecting small targets and adapting to complex backgrounds, making it prone to false positives in multi-target scenarios. YOLOv5, while achieving high detection efficiency through an improved single-stage detection structure and real-time processing of multi-scale targets, still faces limitations in handling small targets, complex backgrounds, and background interference. In maritime target detection tasks, YOLOv5 is particularly sensitive to noise and interference caused by factors such as sea waves and weather changes, which can lead to missed or incorrect detections in complex backgrounds. In contrast, the DTNet network significantly improves robustness to complex backgrounds and noise by adopting specific feature extraction methods (e.g., multi-scale feature fusion) and an innovative detection framework. DTNet is better at capturing the features of maritime targets in different environments, especially in multi-target scenarios, where it effectively distinguishes targets from the background, reducing false positives and missed detections. Additionally, DTNet specifically employs optimization strategies for maritime targets (e.g., handling the interference caused by sea waves and weather changes), allowing it to maintain high precision and stability in complex maritime environments. By optimizing the maritime target dataset, the new model is better equipped to handle challenges such as sea waves, weather changes, and other complex backgrounds. The proposed model is well suited for using navigation radar to detect and track maritime targets in such complex environments, demonstrating improved precision and the ability to handle complex conditions in practical applications, especially when targets and backgrounds are similar, or when targets undergo deformation or occlusion. DTNet's performance is particularly outstanding in these scenarios.

Table 4. Comparison of the evaluation indicators for test results.

Algorithm	Precision (%)	FA (%)	Parameter (M)	FLOPs (GFLOPs)	FPS
Faster R-CNN	86.8	13.2	41.53	180	4.62
RetinaNet	89.3	10.07	34	239	9.87
YOLOv5	91.7	8.3	7.3	118	13.4
DTNet	99.06	0.94	11.3	127	7.64

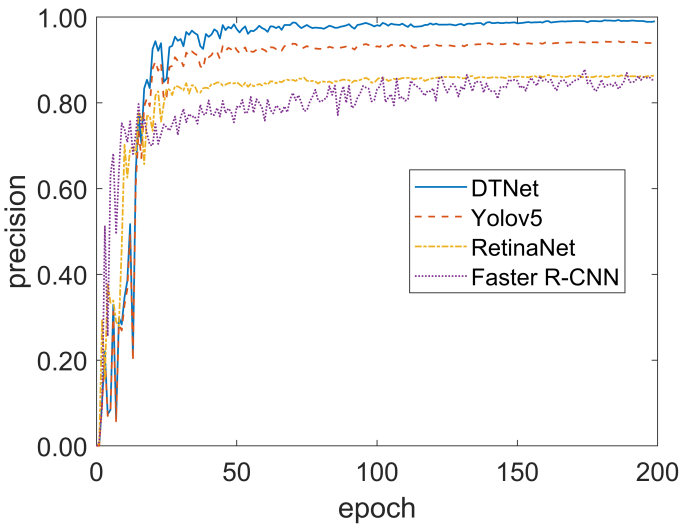


Figure 14. Classic network comparison curve.

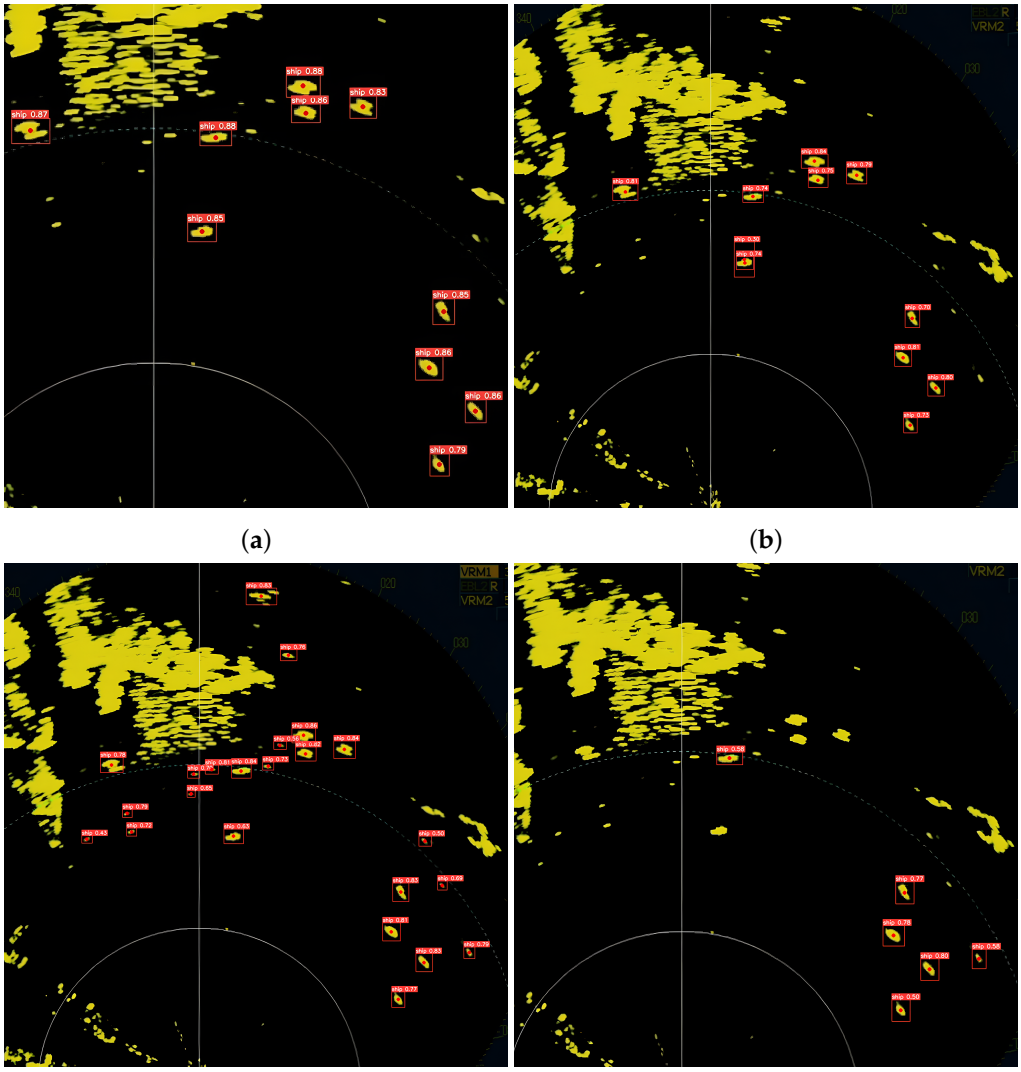


Figure 15. Comparison of the effects with classical algorithm detection. (**a**) DTNet. (**b**) YOLOv5. (**c**) Faster R-CNN. (**d**) RetinaNet.

3.4. Generalization Ability Analysis

3.4.1. Performance Analysis of the Algorithms on Different Maritime Data

(1) The Maritime Area I data were used as the source domain, and they were transferred to the target domain Maritime Area II data.

The model was initially trained using the Maritime Area I dataset, with the weights frozen, before transferring the model to the Maritime Area II dataset for detection and tracking tasks. As shown in Figure 16, the precision steadily improves with an increasing number of iterations, eventually stabilizing at 92.2%, while the loss continuously decreases and stabilizes at 0.08. The white dots marked in Figure 17a represent the AIS positions of targets in Maritime Area II. The detection results are shown in Figure 17b. Compared with AIS, it can be seen that the proposed algorithm correctly detects targets even in the more complex coastal boundary scenarios. The tracking results, as depicted in Figure 17c, show that the DeepSORT tracking algorithm accumulates white tracking trajectories frame by frame. In comparison with the AIS data, it can be observed that the proposed algorithm demonstrates strong tracking performance in Maritime Area II.

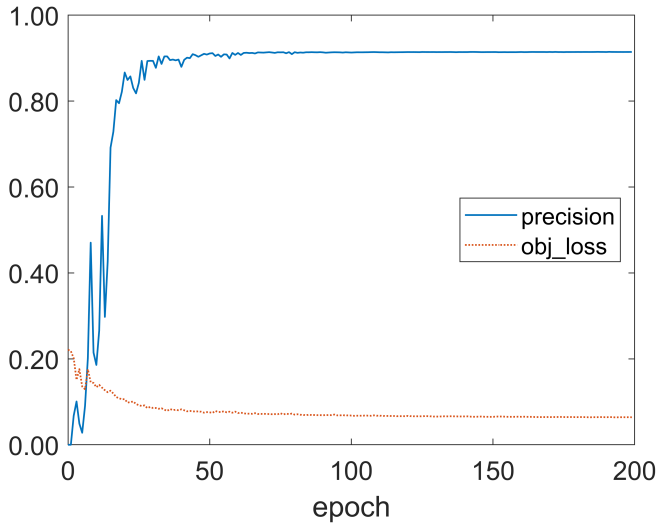


Figure 16. Precision and loss curve for the transfer from Maritime Area I to Maritime Area II.

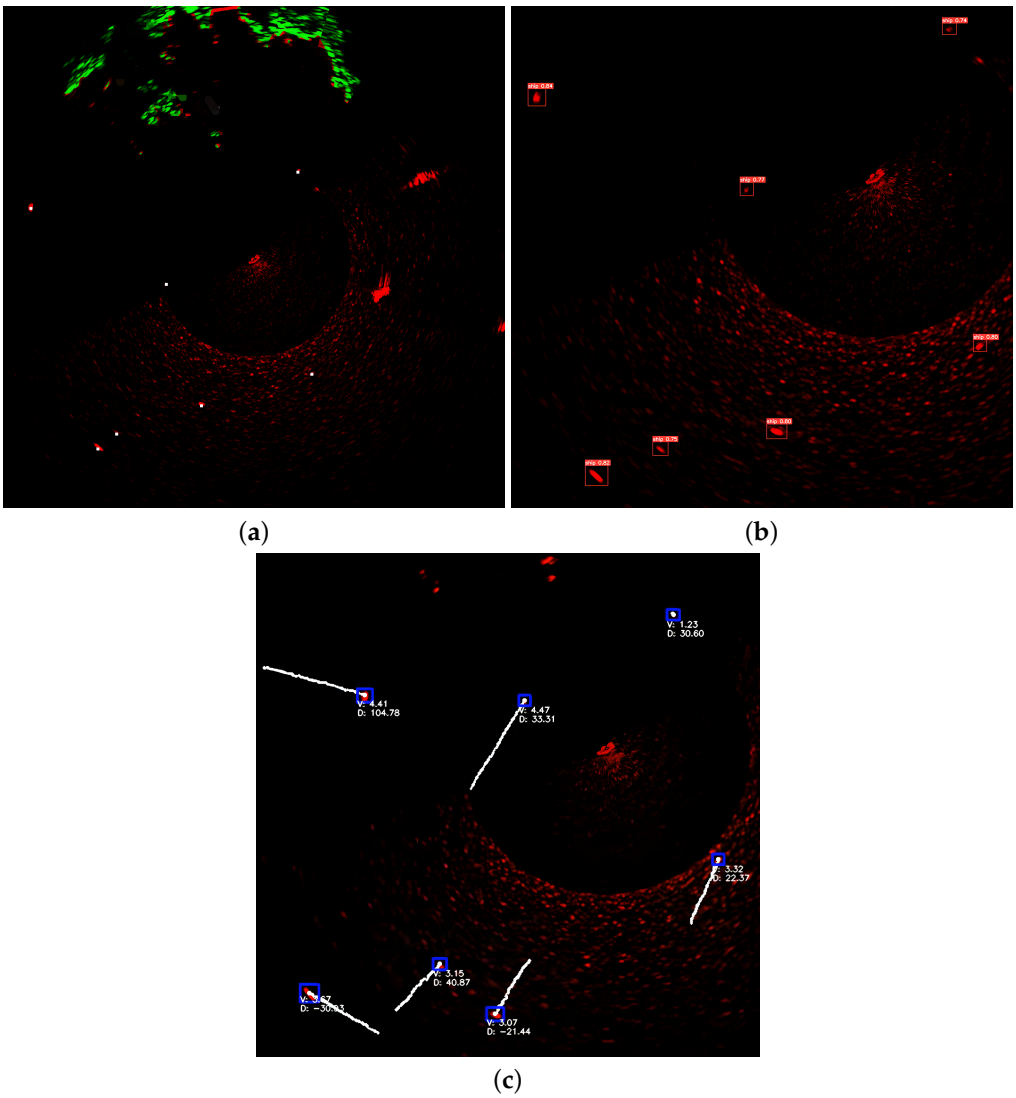


Figure 17. The AIS and detection-tracking results of Maritime Area II. (**a**) The AIS display of the targets in Maritime Area II. (**b**) Detection results of the model transfer from Maritime Area I to Maritime Area II. (**c**) Tracking results of the model transfer from Maritime Area I to Maritime Area II.

(2) The Maritime Area I data is used as the source domain, transferring to the target domain Maritime Area III data.

The model was initially trained using the Maritime Area I dataset, with the weights frozen before transferring the model to the Maritime Area III dataset for detection and tracking tasks. As shown in Figure 18, the precision steadily increases with the number of iterations, eventually stabilizing at around 89%, while the loss continuously decreases, reaching stability at 0.052. The white dots marked in Figure 19a represent the AIS positions of the targets in Maritime Area III. The detection results are shown in Figure 19b. Compared with AIS, it can be seen that the proposed algorithm correctly detects targets in the open sea and achieves a higher detection probability. The tracking results, as shown in Figure 19c, demonstrate that the proposed algorithm performs remarkably well in Maritime Area III, as evidenced by the comparison with the corresponding AIS data.

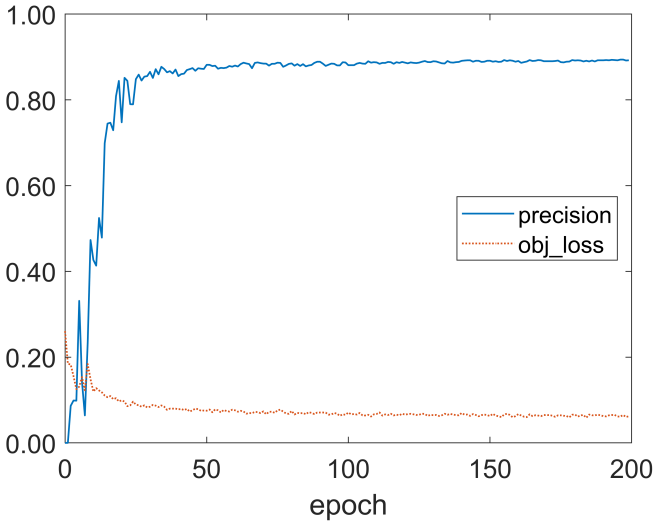
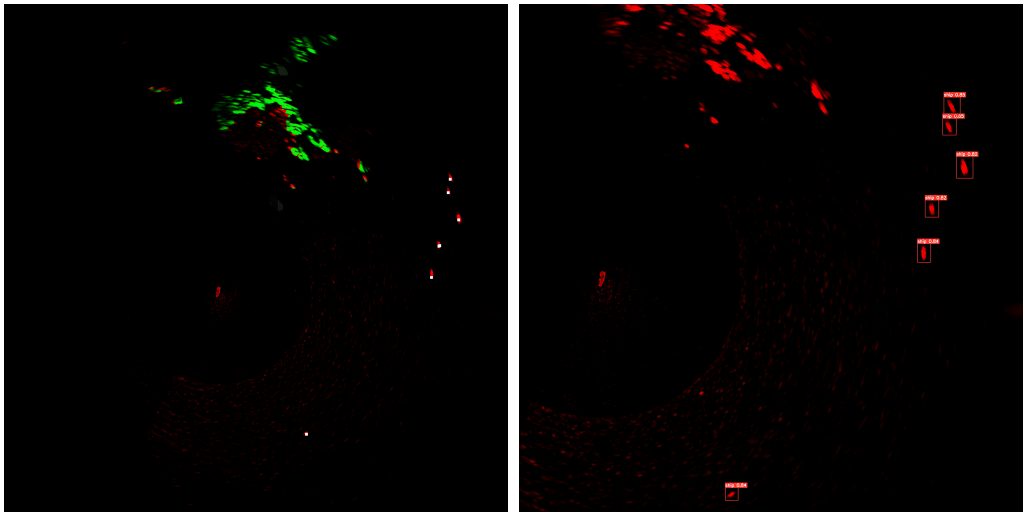
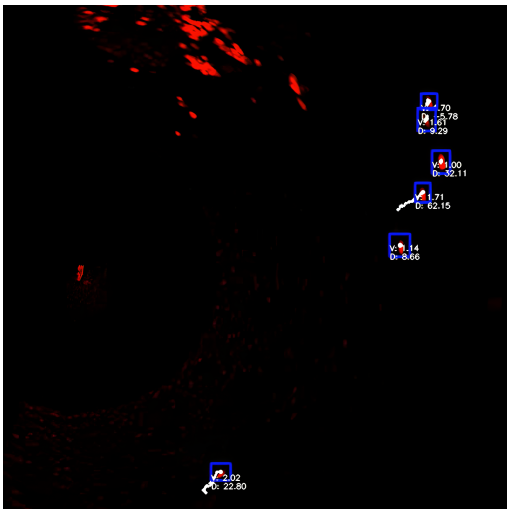


Figure 18. The precision and loss curve for transfer from Maritime Area I to Maritime Area III.



(**a**) (**b**)

Figure 19. *Cont.*



(**c**)

Figure 19. The AIS and detection-tracking results of Maritime Area III. (**a**) The AIS display of the targets in Maritime Area III. (**b**) The detection results of the model transfer from Maritime Area I to Maritime Area III. (**c**) The tracking results of the model transfer from Maritime Area I to Maritime Area III.

3.4.2. Performance Analysis of the Algorithms Under Different Weather Conditions

Since heavy rain conditions involve both sea clutter and rain clutter, this section focuses solely on rainy weather as it is more representative. The model was trained using the clear weather dataset from Maritime Area I, and its weights were frozen before transferring the model to the heavy rain dataset from Maritime Area I for detection and tracking tasks. As shown in Figure 20, the precision gradually increased with the number of iterations, eventually stabilizing at around 82.4%, while the loss decreased steadily, stabilizing at 0.047. In Figure 21a, the red boxes represent the AIS information of the targets. The detection results are shown in Figure 21b. Compared with AIS, it can be seen that the proposed algorithm correctly detected targets, even in heavy rain conditions and with a higher detection probability. This demonstrates that DTNet has strong robustness. The tracking results, as illustrated in Figure 21c, show that, under heavy rain conditions, sea clutter becomes more pronounced, causing greater interference, which leads to reduced precision compared to previous transfers and an increased likelihood of misclassification.

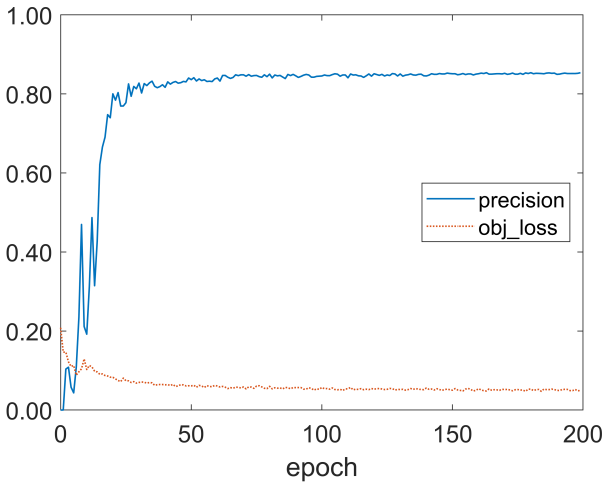


Figure 20. The precision and loss curve for transfer from Maritime Area I to heavy rain in Maritime Area I.

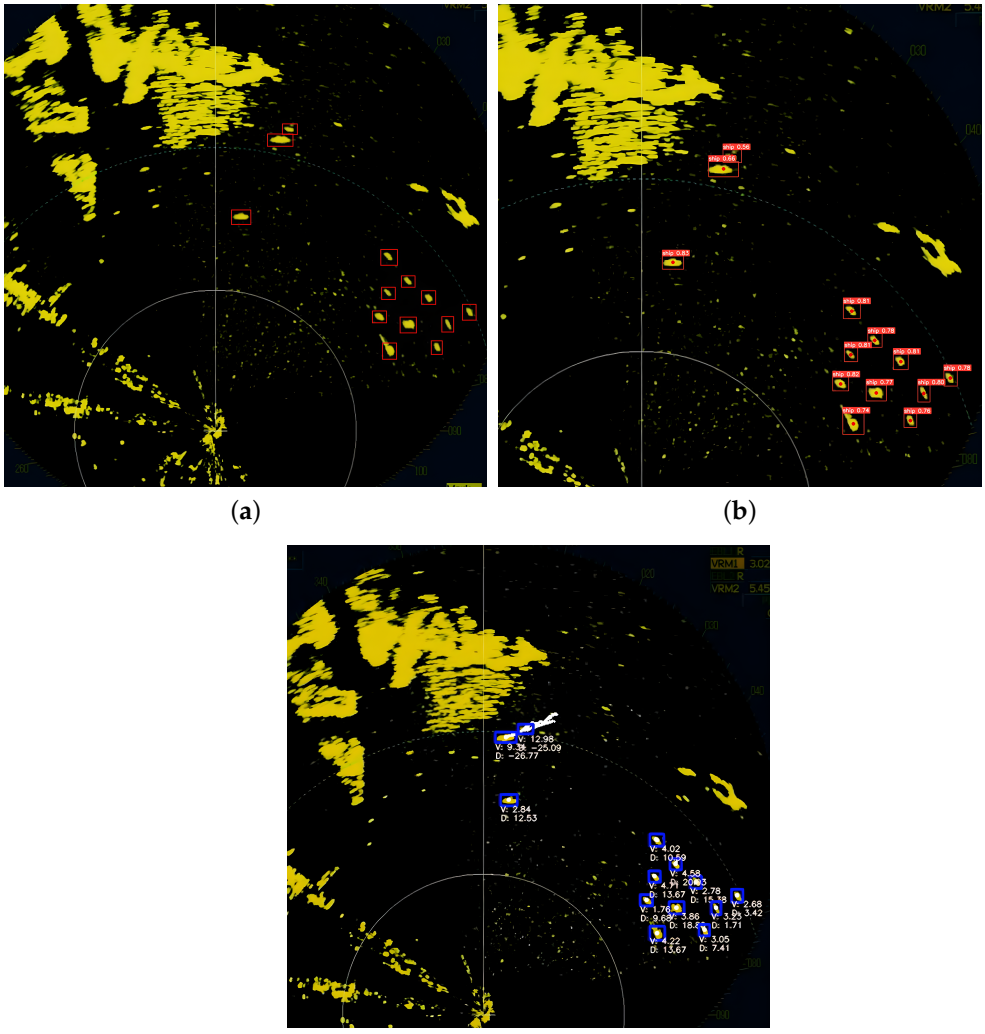


Figure 21. The AIS and detection-tracking results of Maritime Area I during rainy weather. (**a**) The AIS display of the targets in Maritime Area I during heavy rain. (**b**) The detection results for transfer from Maritime Area I to heavy rain in Maritime Area I. (**c**) The tracking results for transfer from Maritime Area I to heavy rain in Maritime Area I.

4. Discussion

The research presented in this paper addresses the critical challenges of maritime target detection and tracking in complex environments. Traditional methods often struggle to generalize across varying maritime conditions due to their inherent limitations in adapting to dynamic scenarios. This study demonstrates that integrating advanced techniques, such as Transformer and the CBAM with the YOLOv5 framework, significantly enhances detection performance.

The proposed model achieves an impressive target detection precision of 99.06%, representing a notable 7.36 percentage point improvement over the original YOLOv5. This advancement can be attributed to the effective incorporation of inter-frame information, which allows the model to differentiate between moving targets and background sea clutter more effectively. The combination of the DeepSORT tracking algorithm further enhances real-time tracking capabilities, making this approach suitable for practical maritime applications.

The findings regarding the generalization capabilities of the model are particularly noteworthy. The ability to transfer learning from Maritime Area I to Areas II and III, with precisions of 92.2% and 89%, respectively, highlights the robustness of the model in different

environments. This suggests that the integrated network can effectively adapt to varying maritime conditions, which is crucial for operational deployments in diverse settings.

The exploration of weather effects on detection precision reveals the challenges posed by adverse conditions, such as heavy rain. Although the model maintains a precision of 82.4% when transferring from Area I to heavy rain conditions, the decline in performance indicates that sea clutter interference remains a significant challenge. This finding underscores the need for further research into methods for mitigating the impact of environmental factors on detection performance.

Building on the current results, future studies should focus on addressing the challenges posed by severe weather conditions. Potential approaches may include developing more sophisticated filtering techniques to enhance target visibility in adverse weather or employing ensemble methods that combine multiple model outputs to improve robustness. Additionally, exploring the integration of additional sensor data, such as AIS or visual inputs, could further enhance detection and tracking precision.

In summary, the proposed DTNet-based approach significantly advances maritime target detection and tracking by leveraging transfer learning and advanced network architecture. The demonstrated generalization capabilities across different maritime regions and conditions provide a strong foundation for future developments. Continued research into enhancing performance under varying environmental conditions will be essential for improving the reliability of maritime surveillance systems and ensuring effective operational outcomes in real-world scenarios.

5. Conclusions

This paper addresses the problem of maritime target detection and tracking and proposes an integrated model for radar frame-to-frame information maritime target detection and tracking based on DTNet transfer learning. Building upon YOLOv5, the Transformer and CBAM modules are introduced to more accurately separate targets from the background by incorporating inter-frame information. Combined with the DeepSORT tracking algorithm, the model achieves real-time, high-precision target detection and tracking. In the experiments, echo data collected in real time from navigation radar in actual maritime environments are pre-processed to construct three datasets that represent different maritime areas and two weather conditions in the PPI format. Based on this dataset, the proposed model is trained, initialized with network weights, and fine-tuned globally. The results show that, in Maritime Area I, the target detection precision reached 99.06%. Compared to Faster R-CNN, RetinaNet, and the most optimal YOLOv5 network, the precision was improved by 7.36%. To verify the generalization ability of the model, the influence of different maritime areas on network performance was further studied. Maritime Area I was used as the source domain for pre-training and transferred to Maritime Area II, achieving a detection precision of 92.2%. When transferring from Maritime Area I to Maritime Area III, the detection precision reached 89%. In addition, this paper also explored the impact of weather on the model. Pretraining was conducted using three weather conditions other than rainy days from Maritime Area I as the source domain. When transferred to the heavy rain conditions in Maritime Area I, the detection accuracy reached 82.4%. Although detection performance slightly decreased under heavy rain conditions due to strong sea clutter interference, the results remained satisfactory. It is worth noting that the proposed algorithm still has certain limitations. In particularly complex maritime environments, especially under adverse weather conditions, large waves, and sea clutter, distinguishing between targets and the background becomes extremely challenging. In such complex backgrounds, DTNet may be affected by interference, making accurate detection and tracking difficult. Future research will focus on addressing the interference caused by adverse

maritime weather conditions in target detection and tracking, offering directions to further enhance the performance of integrated maritime target detection and tracking models.

Author Contributions: Conceptualization, X.H. and X.C.; Data curation, X.H.; Formal analysis, X.D.; Funding acquisition, X.C. and X.D.; Investigation, X.H., X.C., X.D., J.G., S.X. and X.W.; Methodology, X.H., X.C. and X.D.; Project administration, X.H., X.C. and X.D.; Resources, X.C.; Software, X.H.; Supervision, X.C., X.D., J.G. and S.X.; Validation, X.H. and X.C.; Visualization, X.H.; Writing—original draft, X.H. and X.C.; Writing—review and editing, X.H., X.C. and X.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported, in part, by the National Natural Science Foundation of China (under Grant 62222120) and, in part, by the Key Research and Natural Science Foundation of Shandong (under Grant ZR2024JQ003).

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding authors.

Acknowledgments: We would like to thank the anonymous referee for their constructive comments that improved the manuscript. We also appreciate the researchers from all participating institutions involved in this project. Additionally, we would like to express our gratitude to Shanghai Radio and Television Communication Technology Co., Ltd. for providing the radar data.

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

References

1. Shi, Y.L.; Guo, Y.X.; Yao, T.T.; Liu, Z.P. Sea-Surface Small Floating Target Recurrence Plots FAC Classification Based on CNN. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5115713. [CrossRef]
2. Bai, Y.; Chi, W.K.; Xie, B.R.; Zhang, L.; Deng, L.Y.; Mu, W.T. Infrared Ship Detection by Combining Target Extraction and Deep Learning. *Telecommun. Eng.* **2023**, *63*, 193–198.
3. Su, N.Y.; Chen, X.L.; Guan, J.; Huang, Y. Maritime Target Detection Based on Radar Graph Data and Graph Convolutional Network. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 4019705. [CrossRef]
4. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]
5. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, 20–26 June 2005; IEEE: New York, NY, USA, 2005; pp. 886–893.
6. Danelljan, M.; Häger, G.; Khan, F.S.; Felsberg, M. Accurate scale estimation for robust visual tracking. In Proceedings of the British Machine Vision Conference, Nottingham, UK, 1–5 September 2014; BMVA Press: London, UK, 2014.
7. Lukežič, A.; Vojíř, T.; Čehovin, L.; Matas, J.; Kristan, M. Discriminative correlation filter with channel and spatial reliability. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 6309–6318.
8. Danelljan, M.; Häger, G.; Khan, F.S.; Felsberg, M. Discriminative scale space tracking. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1561–1575. [CrossRef] [PubMed]
9. Henriques, J.F.; Caseiro, R.; Martins, P.; Batista, J. High-speed tracking with kernelized correlation filters. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 583–596. [CrossRef] [PubMed]
10. Trunk, G.; George, S.F. Detection of targets in non-Gaussian sea clutter. *IEEE Trans. Aerosp. Electron. Syst.* **1970**, *6*, 620–628. [CrossRef]
11. Rui, S. An improved algorithm for detecting targets in sea spikes. *Fire Control Radar Technol.* **2021**, *50*, 69–76.
12. Chen, X.; Mu, X.; Guan, J.; Liu, N.; Zhou, W. Marine target detection based on marine faster R-CNN for navigation radar PPI images. *Front. Inf. Technol. Electron. Eng.* **2022**, *23*, 630–643. [CrossRef]
13. Mou, X.; Chen, X.; Guan, J.; Dong, Y.; Liu, N. Sea clutter suppression for radar PPI images based on SCS-GAN. *IEEE Geosci. Remote Sens. Lett.* **2021**, *18*, 1886–1890. [CrossRef]
14. Chen, X.; Guan, J.; Mu, X.; Wang, Z.; Liu, N.; Wang, G. Multi-dimensional automatic detection of scanning radar images of marine targets based on radar PPInet. *Remote Sens.* **2021**, *13*, 3856. [CrossRef]
15. Chen, X.; He, X.; Deng, Z.; Guan, J.; Du, X.; Xue, W.; Su, N.; Wang, J. Radar intelligent processing technology and application for weak target. *J. Radars* **2024**, *13*, 501–524. [CrossRef]

16. Chen, X.L.; Su, N.Y.; Huang, Y.; Guan, J. False alarm-controllable radar detection for marine target based on multi-features fusion via CNNs. *IEEE Sens. J.* **2021**, *21*, 9099–9111. [CrossRef]
17. Wang, J.; Li, S. Radar target detection in sea clutter based on CNN with dual perspective attention. *IEEE Geosci. Remote Sens. Lett.* **2022**, *20*, 3500405. [CrossRef]
18. Yan, H.; Chen, C.; Jin, G.D.; Zhang, J.; Wang, X.; Zhu, D. Implementation of a modified faster R-CNN for target detection technology of coastal defense radar. *Remote Sens.* **2021**, *13*, 1703. [CrossRef]
19. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; IEEE: New York, NY, USA, 2016; pp. 779–788.
20. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.; Berg, A.C. SSD: Single shot multibox detector. In Proceedings of the European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 8–16 October 2016; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Springer: Cham, Switzerland, 2016; Volume 9905, pp. 21–37.
21. Girshick, R. Fast R-CNN. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; IEEE: New York, NY, USA, 2015; pp. 1440–1448.
22. Li, X.; Zha, Y.; Zhang, T.; Cui, Z.; Zuo, W.M.; Hou, Z.; Lu, H.C.; Wang, H.Z. Survey of visual object tracking algorithms based on deep learning. *J. Image Graph.* **2019**, *24*, 2057–2080. [CrossRef]
23. Bertinetto, L.; Valmadre, J.; Henriques, J.F.; Vedaldi, A.; Torr, P.H. Fully-convolutional siamese networks for object tracking. In Proceedings of the European Conference on Computer Vision Workshops (ECCVW), Amsterdam, The Netherlands, 8–10 and 15–16 October 2016; Springer: Cham, Switzerland, 2016; pp. 850–865.
24. 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.* **2017**, *39*, 1137–1149. [CrossRef]
25. Li, B.; Yan, J.; Wu, W.; Zhu, Z.; Hu, X. High-performance visual tracking with siamese region proposal network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 8971–8980.
26. Li, B.; Wu, W.; Wang, Q.; Zhang, F.; Xing, J.; Yan, J. SiamRPN++: Evolution of siamese visual tracking with very deep networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 4282–4291.
27. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
28. Li, Y.; Peng, C.; Chen, Y.; Jiao, L.; Zhou, L.; Shang, R. A Deep Learning Method for Change Detection in Synthetic Aperture Radar Images. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 5751–5763. [CrossRef]
29. Wen, B.; Wei, Y.; Lu, Z. Sea Clutter Suppression and Target Detection Algorithm of Marine Radar Image Sequence Based on Spatio-Temporal Domain Joint Filtering. *Entropy* **2022**, *24*, 250. [CrossRef]
30. Jian, L.; Wen, G. Maritime Target Detection and Tracking. In Proceedings of the 2019 IEEE 2nd International Conference on Automation, Electronics and Electrical Engineering (AUTEEE), Shenyang, China, 22–24 November 2019; pp. 309–314.
31. Bairami, S.; Salimi, M.; Mirabbasi, D. A Novel Method for Maximum Power Point Tracking of the Grid-Connected Three-Phase Solar Systems Based on the PV Current Prediction. *Chin. J. Electron.* **2023**, *32*, 353–364. [CrossRef]
32. Zhang, T.; Zhu, K.; Zheng, S.; Niyato, D.; Luong, N.C. Trajectory Design and Power Control for Joint Radar and Communication Enabled Multi-UAV Cooperative Detection Systems. *IEEE Trans. Commun.* **2022**, *71*, 158–172. [CrossRef]
33. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *arXiv* **2017**, arXiv:1706.03762.
34. Lin, T.-Y.; Goyal, P.; Girshick, R.B.; He, K.; Dollar, P. Focal Loss for Dense Object Detection. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2999–3007.
35. Zhang, H.; Wang, Y.; Dayoub, F.; Sunderhauf, N. VarifocalNet: An IoU-Aware Dense Object Detector. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 8514–8523.
36. Yangyi, D.; Kang, H.; Qi, H.; Huang, K. Review of CNN-Transformer Hybrid Model in Computer Vision. *Model. Simul.* **2023**, *12*, 3657–3672.
37. Wojke, N.; Bewley, A.; Paulus, D. Simple Online and Realtime Tracking with a Deep Association Metric. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 3645–3649. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.