Article

Multi-Domain Joint Synthetic Aperture Radar Ship Detection Method Integrating Complex Information with Deep Learning

Chaoyang Tian [1,2], Zongsen Lv [1,2], Fengli Xue [1], Xiayi Wu [1] and Dacheng Liu [1,*]

[1] Department of Space Microwave Remote Sensing System, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100190, China; tianchaoyang22@mails.ucas.ac.cn (C.T.); lvzongsen22@mails.ucas.ac.cn (Z.L.); xuefl@aircas.ac.cn (F.X.); wuxy@aircas.ac.cn (X.W.)

[2] School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China

[*] Correspondence: dcliu@mail.ie.ac.cn

Abstract: With the flourishing development of deep learning, synthetic aperture radar (SAR) ship detection based on this method has been widely applied across various domains. However, most deep-learning-based detection methods currently only use the amplitude information from SAR images. In fact, phase information and time-frequency features can also play a role in ship detection. Additionally, the background noise and the small size of ships also pose challenges to detection. Finally, satellite-based detection requires the model to be lightweight and capable of real-time processing. To address these difficulties, we propose a multi-domain joint SAR ship detection method that integrates complex information with deep learning. Based on the imaging mechanism of line-by-line scanning, we can first confirm the presence of ships within echo returns in the eigen-subspace domain, which can reduce detection time. Benefiting from the complex information of single-look complex (SLC) SAR images, we transform the echo returns containing ships into the time-frequency domain. In the time-frequency domain, ships exhibit distinctive features that are different from noise, without the limitation of size, which is highly advantageous for detection. Therefore, we constructed a time-frequency SAR image dataset (TFSID) using the images in the time-frequency domain, and utilizing the advantages of this dataset, we combined space-to-depth convolution (SPDConv) and Inception depthwise convolution (InceptionDWConv) to propose Efficient SPD-InceptionDWConv (ESIDConv). Using this module as the core, we proposed a lightweight SAR ship detector (LSDet) based on YOLOv5n. The detector achieves a detection accuracy of 99.5 with only 0.3 M parameters and 1.2 G operations on the dataset. Extensive experiments on different datasets demonstrated the superiority and effectiveness of our proposed method.

Keywords: synthetic aperture radar (SAR); single-look complex (SLC); complex information; deep learning; ship detection

1. Introduction

Target detection technology for surface ships at sea is significant in both civilian and military fields. Presently, the monitoring task of marine ships mainly relies on various remote sensing platforms. Among these remote sensing platforms, synthetic aperture radar (SAR) satellites have become leading platforms for ship detection because of their advantages of all-weather and all-day characteristics, multi-polarization, strong penetrability, and extensive coverage area [1]. With a large number of SAR ship detection datasets being made public, the ship detection field has entered a phase where deep-learning-based detection has become the mainstream [2], and the detection accuracy obtains substantial improvements. However, compared to natural images, SAR images are low-resolution grayscale images lacking color and detailed information while being heavily affected by

noise. These factors result in poor-quality SAR images, thus limiting the further improvement of detection accuracy in SAR ship detection. Therefore, it is still a challenge to enhance the performance of SAR ship detection [3,4].

Prior to the emergence of deep learning, ship detection using SAR images primarily relied on statistical-model-based methods [5,6], with the most renowned being the constant false alarm rate (CFAR) method and its improvements [7,8]. Benefiting from the prior information that the scattering intensity of ship targets made of metal is higher than that of sea clutter, such methods statistically model the sea clutter and then segment targets from the background based on the derived false alarm probability threshold [9,10]. However, the maritime environment is highly dynamic, making it difficult to accurately model sea clutter. As a result, the robustness of this method is relatively poor, limiting its detection performance. Nowadays, it primarily serves as an auxiliary tool [11,12] for deep-learning-based methods.

With the continuous development of deep learning, its influence has spread to various fields, benefiting SAR ship detection greatly as well. Deep-learning-based object detection techniques have evolved to the present day and can be roughly divided into two architectures: convolutional neural network (CNN)-based and transformer-based. Based on whether anchor boxes are predefined, CNN architectures can be further divided into anchor-based and anchor-free methods. Among them, anchor-based methods can be further subdivided into two-stage and one-stage models. Two-stage models, represented by the RCNN series [13–15], are known for their high accuracy but slow speed and large size, while single-stage models [16] are best known for the YOLO series of models [17,18], which have real-time and lightweight characteristics as their main advantages but are slightly less accurate. Anchor-free methods [19,20] eliminate the constraints of the preset anchor, further improving detection efficiency. Models based on transformer architectures have larger receptive fields [21,22], which helps enhance detection performance. SAR ship detectors are ultimately to be deployed on satellites, which need to have highly lightweight and real-time properties, and single-stage detectors are the most suitable option, considering the accuracy and complexity.

Based on the above analysis, in recent years, researchers have developed numerous SAR ship detectors based on single-stage detectors. Wang et al. introduced data augmentation and transfer learning to propose an enhanced SSD model for SAR target detection [23]. Yang et al. combined the advantages of RetinaNet and the rotatable bounding box and proposed an improved one-stage detector to solve problems such as the unbalanced distribution of positive samples [24]. Sun et al. noted the ship's angle problem and designed an enhanced YOLO-based detector called BIFA-YOLO [25]. Inspired by MobileNet, Yang et al. proposed a super-lightweight and efficient SAR ship detector based on the YOLOv5n [26]. Using YOLOv7 as foundational network, Tang et al. proposed DBW-YOLO to cope with the difficulties that exist in ship detection in complex environments [27]. Zhou et al. introduced a transformer to YOLOX and designed a lightweight SAR ship detection network [28]. Zhou et al. improved the pooling method and loss function of YOLOv5 to enhance the detection accuracy of small ships [29]. On this basis, Liu et al. introduced explainable evidence learning to address the problem of intraclass imbalance [30].

Although the methods mentioned above have achieved excellent performance, they only use the amplitude information of SAR images. This is because phase information is generally considered to be useless. However, in recent years, researchers have found through studies that phase or complex information can play a role in discriminating ships [31]. El-Darymli et al. constructed a statistical model in the complex domain of SAR images, which achieve higher performance [32]. Leng et al. utilized Complex Signal Kurtosis (CSK) to detect ships in complex-valued SAR images, reducing false alarm rates by incorporating complex information [33]. Lv et al. proposed a two-step ship detector of SAR imagery based on complex information to improve detection performance [34].

In the above studies, deep-learning-based methods did not take advantage of the complex information of the SAR images, and the images made with only the amplitude

information suffered from poor image quality problems such as low resolution, grayscale range, noise, and small-sized ships. Small target detection has always been one of the challenging aspects of object detection. Firstly, as the network depth increases, small targets may lack sufficient information to support accurate detection. Secondly, the information of small targets may be overshadowed by that of larger targets, leading to their being overlooked. Lastly, small targets might even be treated as noise and filtered out. These factors, among others, significantly affect detection accuracy. Noise can indeed originate from any of the factors mentioned. Given that SAR images are grayscale and often have a blurred appearance, the noise might resemble the visual characteristics of ships, which can mislead the detector into making incorrect judgments. In order to solve these problems, most detectors have to pay a price in terms of complexity, while complex information-based methods mostly rely on hand-crafted rules, which results in their low robustness.

To tackle these challenges, we propose a multi-domain joint SAR ship detection method that combines the merits of complex information and deep learning. We first use eigen-subspace projection (ESP) techniques to map echo returns of the single-look complex (SLC) SAR image to the eigen-subspace domain to determine the presence of ships. This step filters out a large number of irrelevant echoes, significantly speeding up the overall detection process and ensuring real-time performance. Then, the echo returns containing ships are transformed into the time-frequency domain, which dramatically reduces the detection difficulty. In the time-frequency domain, we detect the ship using a deep-learning-based method and then transform the image back to the image domain using the phase information to locate the actual position of the ships. In this step, we leveraged the advantages of the data to design a lightweight and efficient model, which further accelerates the detection speed while ensuring the model's compactness.

The primary contributions of this article are as follows:

(1) This article proposes a multi-domain joint SAR ship detection method combining complex information and deep learning, which exploits the characteristics of ships in the eigen-subspace domain and time-frequency domain within SAR SLC data, combined with the advantages of deep learning, to significantly improve detection efficiency.

(2) To meet the training requirements of deep learning, we constructed a dataset that contains 640 amplitude images of SLC SAR imagery in the time-frequency domain, which was divided into training and test sets in the ratio of 8:2.

(3) Taking advantage of the dataset, a lightweight ship detector was designed. The detector deletes the detection neck, which is widely adopted in single-stage detectors for fusing the multi-scale features, and achieves a better balance between computational complexity and accuracy.

2. Proposed Method

The flowchart of the proposed method is shown in Figure 1. The whole method can be roughly divided into two steps. The first step is to judge whether the ship exists through the ESP technique [35,36]. We then scan the SLC SAR data line by line, construct the Hankel structure and covariance matrix in turn, carry out the eigenvalue decomposition (EVD) processing to find the maximum eigenvalue, and finally obtain the sequence of maximum eigenvalues. Then, we perform the threshold segmentation of the sequence, and the larger-than-threshold value is determined to be the existence of the ship. The second step is to locate the position of the ship by deep learning method, after confirming the existence of the ship, we first obtain its amplitude and phase in the time-frequency domain by the short-time Fourier transform (STFT) [37], then detect the ship in the amplitude and merge it with its corresponding phase, and finally transform it back to the image domain to determine the actual position of the ship.

2.1. Assessing the Existence of Ships

The scattering characteristics of the ship are much stronger than those of the sea clutter, which can be found from Figure 2. In the figure, the ship appears prominently with a bright

illumination, while the sea clutter is subdued and unilluminated. We select a signal of interest (SOI) containing the ship target in the azimuth direction for in-depth observation, and the results are shown in Figure 3, from which it is obvious that the scattering intensity of the sea clutter is lower than that of the ship. By using ESP technology, we project the signal containing the ship target into the eigen-subspace and find that the strong scattering characteristics of the ship can also be characterized. As shown in Figure 4, the eigenvalues of the ship are much higher than those of the sea clutter.
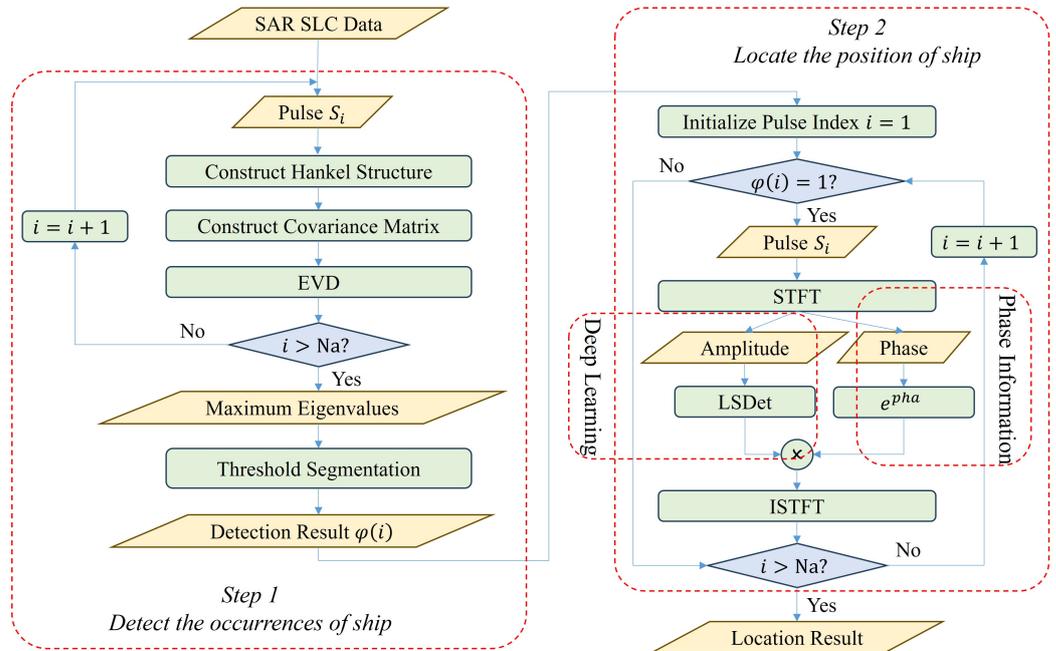

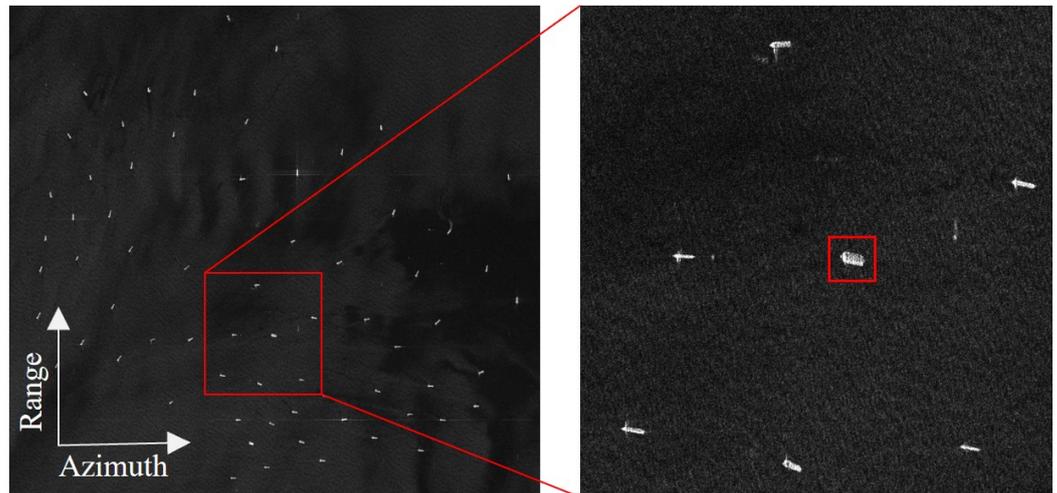**Figure 1.** Flowchart of the proposed method.

Figure 2. Strong scattering characteristics of ships in the image domain.

ESP decomposes the autocorrelation matrix of the observed signal to obtain its eigenvalues by EVD, then construct the subspace of the signal of interest according to the distribution of its eigenvalues, and finally projects the SOI by using the orthogonality between the subspaces. This technique is mostly used in the fields of sparse signal estimation, the direction of arrival, and so on. Based on this, we apply this technique to the processing of SAR signals, with the following specific procedures.
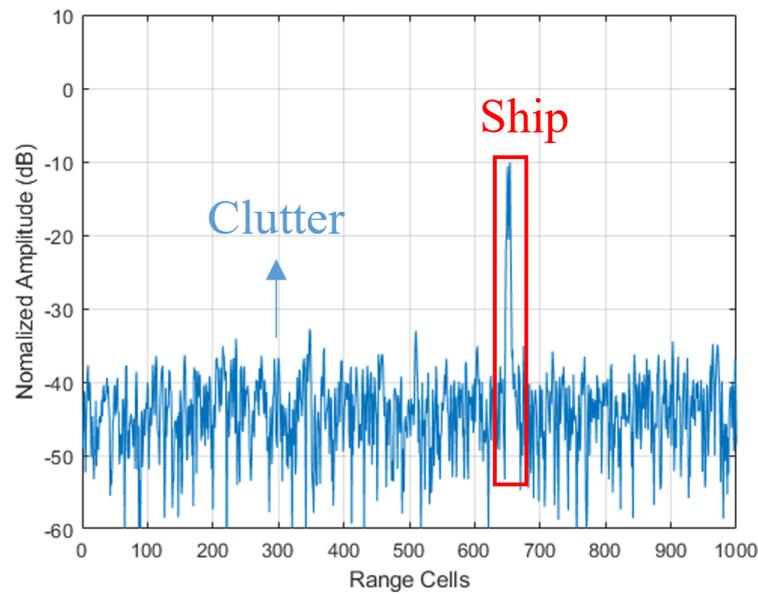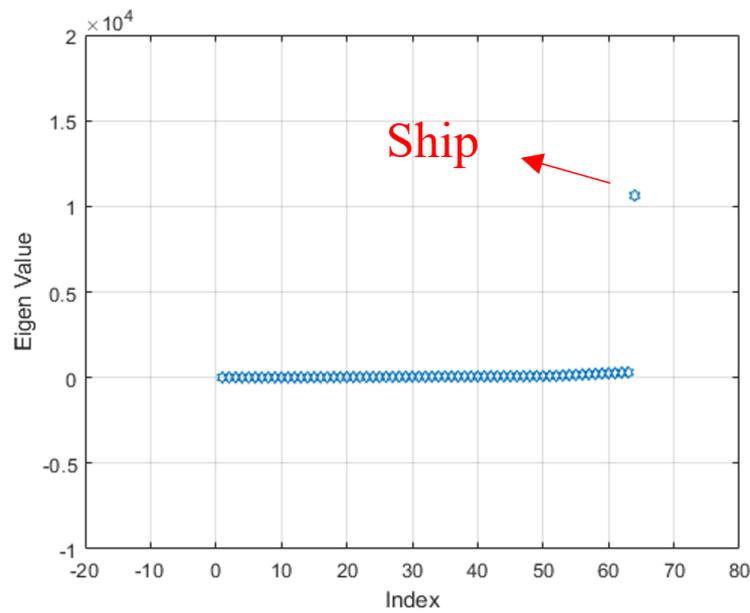
Figure 3. Pulse containing a ship in Figure 2.

Figure 4. Eigenvalue of the Pulse in Figure 3.

The SAR SLC data can be denoted by $S_{sc} = \left[S_{sc}^1, S_{sc}^2, \ldots, S_{sc}^{N_a} \right]^T$ and $S_{sc}^n(\tau) \in \mathbb{C}^{1 \times N_r}$, $\tau = 1, 2, \ldots, N_r$. Based on the Hankel structure, the subspace matrix of S_{sc} can be represented as

$$H_n = \begin{bmatrix} S_{sc}^n(1) & S_{sc}^n(2) & \cdots & S_{sc}^n(M) \\ S_{sc}^n(2) & S_{sc}^n(3) & \cdots & S_{sc}^n(M+1) \\ \vdots & \vdots & \ddots & \vdots \\ S_{sc}^n(L) & S_{sc}^n(L+1) & \cdots & S_{sc}^n(M+L-1) \end{bmatrix} \tag{1}$$

where L is the dimension of the subspace matrix, $M = N_r + 1 - L$ and $n = 1, 2 \ldots, N_a$. N_a and N_r represent the number of samples in the azimuth and range directions, respectively. These variables indicate the resolution of the signal sampling rate in different directions.

From the matrix H_n, the covariance matrix X_n can be constructed

$$X_n = H_n H_n^T \tag{2}$$

Then, we decompose X_n by the EVD

$$X_n = E_n \Lambda_n H_n^T \tag{3}$$

where $\Lambda_n = diag(\lambda_1^n, \lambda_2^n, \lambda_3^n, \ldots, \lambda_L^n)$ and λ^n is the eigenvalue. Repeating the above process, we can obtain the maximum eigenvalue sequence, which can be expressed as

$$\Phi(n) = [\phi_1, \phi_2, \phi_3, \ldots, \phi_{N_a}] \tag{4}$$

where $\phi_n = max[\lambda_1^n, \lambda_2^n, \lambda_3^n, \ldots, \lambda_L^n]$.

Finally, threshold segmentation is performed

$$\hat{\Phi}(n) = \begin{cases} 0 & \Phi(n) < \theta \\ 1 & \Phi(n) \geq \theta \end{cases} \tag{5}$$

where $\hat{\Phi}(n)$ is the judgment result, and θ is the threshold value, derived from statistical analysis of a large amount of experimental data. If $\hat{\Phi}(n)$ is not less than this value, it indicates the presence of a ship.

2.2. Detecting the Positions of Ships
2.2.1. STFT

SAR images are produced by satellites moving in the azimuth direction to gather data along the range direction, synthesizing 1D signals into 2D images. Benefiting from this unique imaging mechanism, each azimuthal echo is essentially a 1D signal. We perform STFT analysis along the azimuth direction on the signals containing the ship, transforming them into the time-frequency domain to determine the exact position of the ship. The comparison of ships in different domains is shown in Figure 5. From the figure, we can easily observe that the features of the ship are more abundant in the time-frequency domain, indicative of a reduced complexity in detection, which is the primary motivation for our detection based on the time-frequency domain.
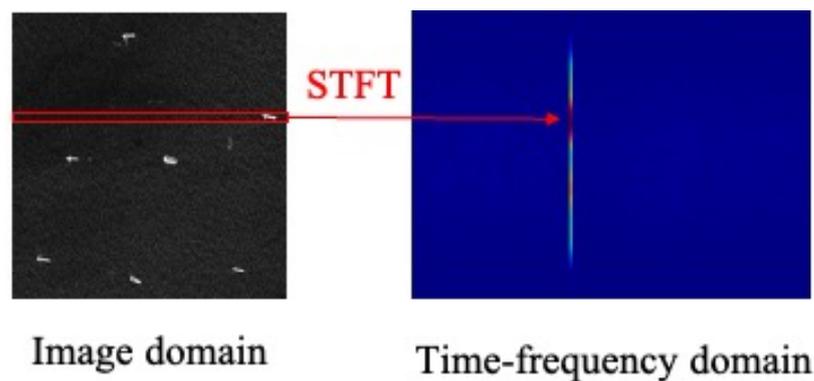
Figure 5. Comparison of ships in different domains.

STFT is a variant of Fourier transform (FT), which solves the problem of handling non-stationary signals encountered in traditional Fourier analysis. FT can represent the signal in the frequency domain, but it is not able to represent the change in the signal over time until the emergence of STFT. STFT is improved by multiplying the signal with a time-limited window function before FT while assuming that the signal is stable within the window. Moving the

window function, the signal is segmented and analyzed to obtain the local spectrum of the signal. The discrete STFT spectrum of the signal $s(p)$ can be expressed as

$$STFT(p,q) = \sum_{u} h(u-p)s(u)exp(-j\frac{2\pi}{N_r}qu) \qquad (6)$$

where p and q represent the number of samples in time and frequency, respectively, and u denotes the signal integer index.

2.2.2. TFSID

Deep-learning-based methods rely on vast quantities of accurately labeled data to show excellent performance. Therefore, to fulfill this need, we constructed a time-frequency domain dataset using images called TFSID captured by the Sentinel-1 satellite. The original images of this dataset consist of six SLC SAR images taken at the Port of Santos in Brazil, all with an image size of 25,689 × 12,458, and are shown in Figure 6. The image of a ship in the time-frequency domain is not related to the location of the original image but rather to the environment in which the ship is located. This is because the ship's representation in the time-frequency domain is entirely determined by its time-frequency characteristics, which are affected by the surrounding environment rather than its spatial position. Thus, the original six images were taken at the same location but in different sea state conditions.
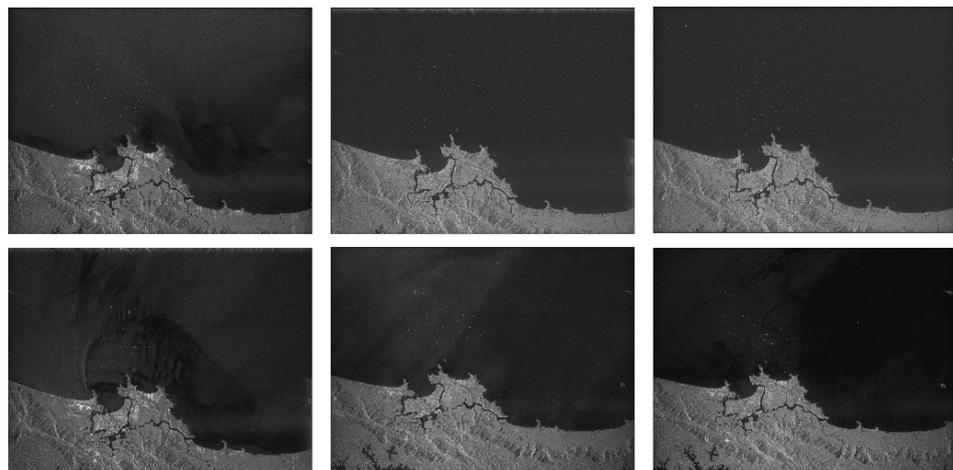

Figure 6. The original images of TFSID.

The images required for model training should contain the ship, so we first observe the approximate position where the ship is located, then delineate a range accordingly, and transform the azimuthal direction cells of this range to the time-frequency domain through SIFT, and finally about 6000 time-frequency images, each with a resolution of 875 × 656 pixels, can be generated from these six original images. Still, not all of these images contained ships, so we conducted a preliminary screening, with images containing only sea clutter being discarded and around 3000 images being retained. The ship features in the time-frequency domain were very distinct, and 3000 images were redundant, so we further filtered them by removing highly similar images and blurred images, and finally 640 images were selected to form the dataset called TFSID. We accurately annotated the partial images of the dataset, which can be seen in Figure 7. A total of 512 images from the dataset were used as training set, and the remaining images were reserved for testing.

When selecting the images, we adhered to strict deduplication and quality control standards to ensure that the dataset was both representative and diverse. Although 640 images may seem like a small number, the images in TFSID possess extremely clear features, making this quantity sufficient for the model to learn and accurately detect characteristics. Moreover, one of the challenges in applying deep learning is the need for large amounts of

well-annotated data to achieve good performance. Our approach overcomes this challenge by training a high-performing model with only a small number of well-annotated images.
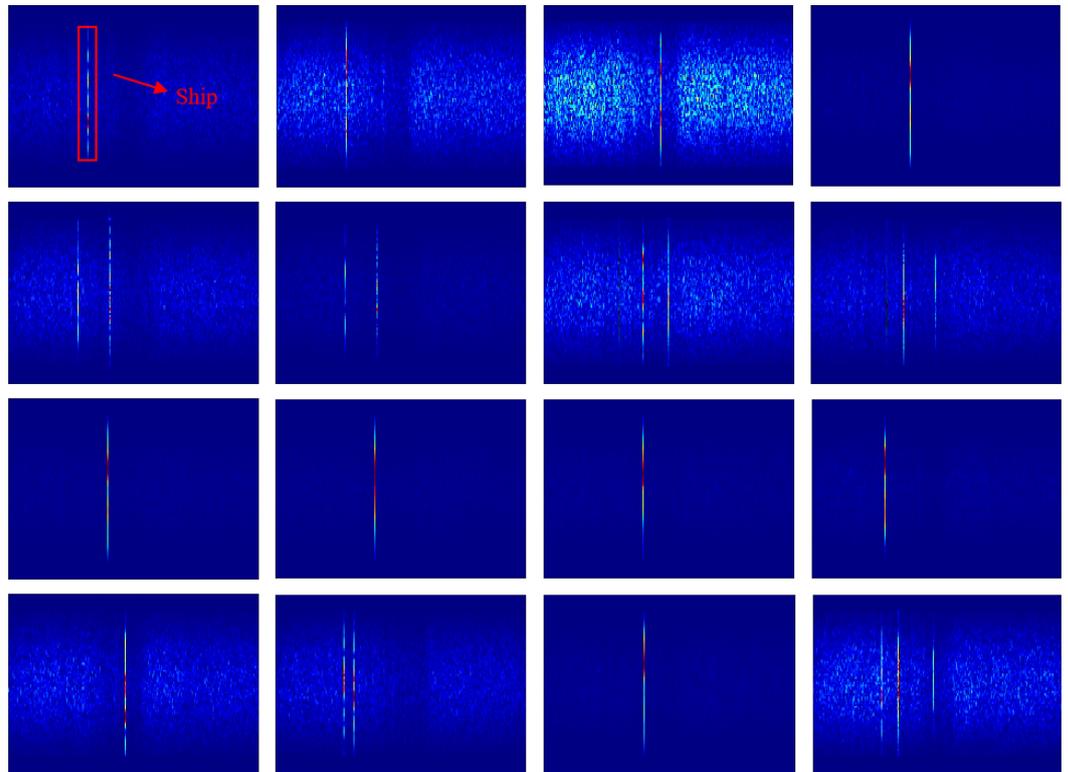
Figure 7. The partial images of TFSID.

2.2.3. LSDet

From Figure 7, it is evident that ships are highly prominent in the time-frequency domain. Therefore, we selected the smallest-sized model, YOLOv5n, from the single-stage detectors as the base model. YOLOv5n uses strided convolution for downsampling; however, recent research explains that this approach has certain drawbacks that can lead to loss of detailed information and poorly learned features [38,39]. Moreover, the recent success of transformers has demonstrated limitations in CNN architectures due to their finite receptive fields, which restrict detection performance. Although this issue can be addressed by increasing the convolutional kernel size, the resultant increase in complexity is deemed unacceptable.

To address the aforementioned issues, inspired by the lightweight philosophy of the MobileNet series models, we designed efficient SPD-InceptionDWConv (ESIDConv), which integrates inception depthwise convolution (InceptionDWConv) [40] and space-to-depth convolution (SPDConv). The specific construction of ESIDConv can be observed from Figure 8. Given a feature map X with dimensions $H \times W \times C$, it undergoes the following process within ESIDConv. Firstly, the feature map is divided into four sub-feature maps of size $H/2 \times W/2 \times C$ along the spatial dimensions as follows:

$$
\begin{aligned}
f_{0,0} &= X[0:2:W, 0:2:H]; \\
f_{1,0} &= X[1:2:W, 0:2:H]; \\
f_{0,1} &= X[0:2:H, 1:2:W]; \\
f_{1,1} &= X[1:2:H, 1:2:W].
\end{aligned}
\tag{7}
$$

Next, each of the four sub-feature maps undergoes a different operation: depthwise separable convolution with a 3×3 kernel (DWConv 3×3), DWConv 1×11, DWConv 11×1, and Identity, respectively. This can be regarded as a variant of the Inception

structure. Then, we concatenated the four refined sub-feature maps along the channel dimension to obtain an intermediate feature map X_i of size $H/2 \times W/2 \times 4C$, which represents the transition from space to depth. Finally, we use a non-strided convolution with a 1×1 kernel to resize its channel dimension back to C, obtaining an output feature map of size $H/2 \times W/2 \times C$. Thus, we efficiently and effectively completed downsampling and feature extraction. By changing the downsampling method and introducing parallel large-kernel DWConv, we successfully circumvented some inherent shortcomings of CNN architectures and achieved a superior trade-off between accuracy and complexity. This implies that we can not only efficiently leverage the advantages of time-frequency domain data but also meet the stringent requirements of lightweight design and real-time performance for satellite-based detection.
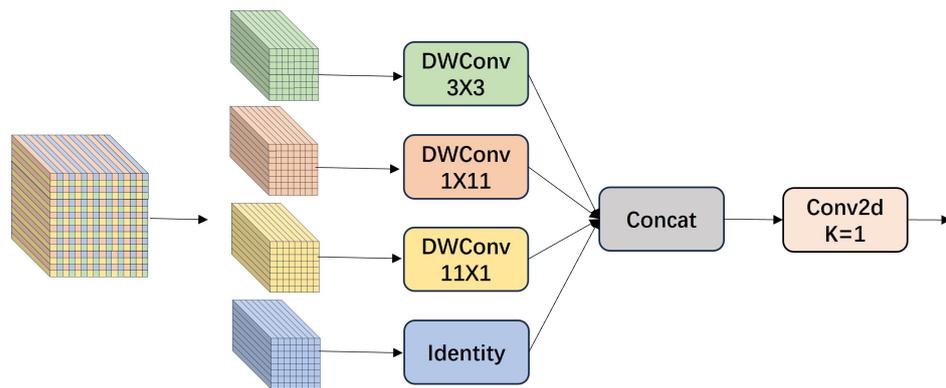

Figure 8. Structure of the ESIDConv.

With ESIDConv as the core module, we propose a new backbone, replacing the original CSPDarknet53 in YOLOv5n. Traditional detection architectures typically append a neck structure after the Backbone, which aims to address the variability in object sizes by fusing multi-scale features. After careful observation, we discovered that the characteristic dimensions of SAR ships in the time-frequency domain are essentially consistent, eliminating the need for a neck structure. Therefore, we can entirely discard the neck component to make the model more lightweight and efficient. Based on the above analysis, we propose a lightweight SAR ship detector named LSDet, whose detailed information can be found in Figure 9.
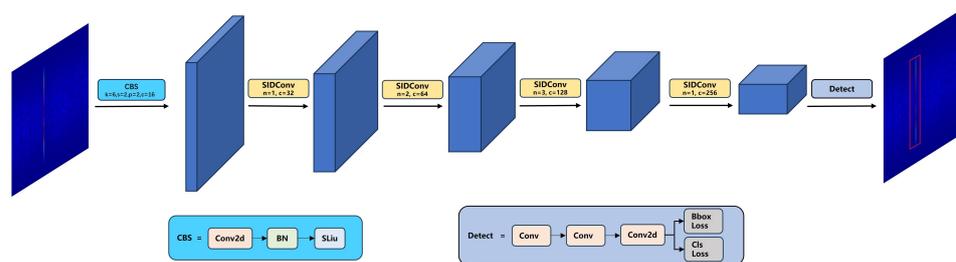
Figure 9. Overall architecture of the proposed LSDet.

After transforming SLC SAR ship images into the time-frequency domain using SIFT, both the amplitude and phase information of the ships can be obtained. We feed the amplitude information into LSDet to detect the ships then merge the corresponding phase information. Finally, by applying inverse SIFT (ISIFT), we transform them back into the image domain to reveal the specific positions of the ships.

3. Experiments

3.1. Experimental Settings

All experiments were performed on a consistent computer setup. The hardware configuration included an RTX 3050 GPU (manufactured by NVIDIA, Santa Clara, CA,

USA) with 4 GB VRAM and a 12th Gen Intel Core i5-12500H processor clocked at 3.10 GHz. The software environment features Python 3.8, CUDA 11.1, and PyTorch 1.8. The system was run on Windows 11.

For training, we did not use any pre-training files. To ensure the fairness and generalizability of the experiments, all relevant parameters were kept as YOLOv5n's default settings, except for the batch size, which was adjusted according to the GPU memory. These settings included using the Adam optimizer, 300 epochs, an input image size of 640×640, and an optimizer weight decay of 0.0005, among others.

3.2. Datasets

To fully verify the efficiency and superiority of our proposed method, we performed comparative analysis experiments on two different domains. In the image domain, we choose the simplest SAR ship detection dataset (SSDD) [41]. The SSDD dataset contains 1160 images, featuring a total of 2456 ships. Twenty percent of the images, specifically those with filenames ending in 1 or 9, were designated as the test set, while the rest were utilized for training and validation. The images were primarily sourced from RadarSat-2, TerraSAR-X, and Sentinel-1 sensors, with resolutions varying between 1 m and 15 m, and each image had a size of 500×500 pixels. In the transforming domain, we used self-built TFAID to carry out experiments.

3.3. Evaluation Indicators

When evaluating the effectiveness of our proposed method, three key performance metrics were introduced: precision, recall, and average precision (AP). These metrics consist of four components: true positives (TPs), false positives (FPs), true negatives (TNs), and false negatives (FNs). TPs refer to correctly predicted positive samples (ships), while FPs correspond to incorrectly predicted positive samples. TNs represent correctly predicted negative samples (background), and FNs indicate incorrectly predicted negative samples. Precision, denoted as P, represents the proportion of accurately predicted ships. Recall, denoted as R, represents the total number of ships retrieved. These two metrics can be derived from the following equations

$$P = \frac{TP}{TP + FP}, R = \frac{TP}{TP + FN} \tag{8}$$

Neither P nor R alone can comprehensively evaluate the model performance. Therefore, AP is introduced, which can be expressed as

$$AP = \int_0^1 P(R) \cdot dR \tag{9}$$

AP at IoU = $0.5(AP_{50})$ and at IoU = $0.75(AP_{75})$ were used to validate the detection accuracy of the model. Additionally, we used FLOPs, parameters, and frames per second (FPS) to assess the model's size and detection speed.

3.4. Experimental Comparison and Analysis

To validate the effectiveness of the proposed module and strategies, we conducted ablation experiments using YOLOv5n as the baseline for the TFSID dataset. The experimental results are presented in Table 1.

As seen from Table 1, neither using the Backbone composed of ESDIConv nor removing the neck part reduces the core detection accuracy (AP_{50}). Instead, both approaches significantly reduce the parameters and FLOPs to about 60% of the original while accelerating the detection process with an approximately 25% increase in FPS. More importantly, when both strategies are applied simultaneously, the detection accuracy remains stable, while the parameters and FLOPs can be further reduced to about 20% of the original,

and the FPS is nearly doubled. These observations strongly confirm the effectiveness of our proposed methods.

Table 1. Ablation experiment on the TFSID dataset.

ESIDConv	Without Neck	AP_{50}	Parameter	FLOPs	FPS
×	×	**99.5**	1.7 M	4.1 G	126.4
×	✓	**99.5**	1.1 M	2.9 G	156.3
✓	×	**99.5**	1.0 M	2.7 G	161.2
✓	✓	**99.5**	**0.3 M**	**1.2 G**	**215.9**

To prove the superiority of our proposed method, we first conducted comparative experiments on TFSID with various representative detectors as well as the detector we designed. These detectors include: the swin transformer-tiny (swin-T) based on the transformer structure [42], the FCOS method based on the anchor-free approach [43], the two-stage model mask R-CNN [44], and the single-stage models [45] and YOLO series [46]. The experimental results are presented in Table 2.

Table 2. Comparison analysis of different types of detectors and our proposed detector for the TFSID dataset.

Method	AP_{50}	AP_{75}	Parameter	FLOPs	FPS
Swin-T	98.1	63.4	48 M	267 G	11.9
FCOS	98.8	71.2	86 M	745 G	56.6
Mask-RCNN	99.1	67.4	86 M	197 G	23.5
retinaNet	98.5	63.4	9 M	26.8 G	30.5
YOLOv5n	**99.5**	72.6	1.7 M	4.1 G	126.4
YOLOv7-tiny	99.3	**77.9**	6 M	13 G	128.5
LSDet (ours)	**99.5**	57.9	**0.3 M**	**1.2 G**	**215.9**

As shown in Table 2, our model has 0.3 M parameters, which is only about 20% of the size of the lightest model YOLOv5n among other models. Despite this, we still achieve the same AP_{50} accuracy as YOLOv5n, both nearing the limit of 99.5. LSDet's FLOPs are also lower than those of other models by over 50%, indicating significantly lower computational complexity. Importantly, our model achieves the fastest detection speed, reaching 215.9 FPS. However, our designed model performs the worst in terms of AP_{75} accuracy, which may be due to the removal of the neck component, resulting in less precise localization. Nonetheless, this impact is minimal, as AP_{50} is already sufficient for localization needs. Overall, the transformer-based model, swin-T, performs the worst. This is likely because the small dataset cannot satisfy the data requirements for training transformers. Taken together, our model performs the best overall, which fully demonstrates the validity of our proposed method and the efficiency of our designed detector.

To further validate the effectiveness of our proposed method of transforming into the time-frequency domain, we conducted additional experiments in the image domain. The SAR ship detection dataset SSDD [41], which is the simplest public dataset in the image domain, was used for the experiments. In addition to various types of SOTA models, we also introduced models [47,48] specifically designed for SAR ship detection in the image domain to ensure the comprehensiveness of the experiments. Table 3 displays the experimental results.

From Table 3, it can be seen that even on the simplest dataset in the image domain, LSDet, a lightweight model, does not achieve the detection performance as in the time-frequency domain. However, due to LSDet's choice of the downsampling method and the introduction of parallel large-kernel DWConv, it partially alleviates the issues of low resolution, dense arrangement, and small ship sizes in SAR ship images in the image domain.

Therefore, its detection performance still surpasses that of the first three types of detectors. Despite this, due to the lack of a neck component, LSDet's AP_{50} and AP_{75} accuracy is lower than that of YOLOv5n, and it cannot be compared to methods specifically designed for SAR ship detection in the image domain. The above results strongly demonstrate the necessity and correctness of our proposed method. Specifically, the method of transforming SLC SAR images into the time-frequency domain to overcome detection difficulties in the image domain has been proven to be both feasible and highly efficient.

Table 3. Comparison of different methods for the SSDD datasets.

| Method	AP_{50}	AP_{75}	Parameter	FLOPs	FPS
Swin-T	94.6	68.3	48 M	267 G	10.1
Mask-RCNN	91.3	63.4	86 M	197 G	17
FCOS	93.2	67.8	32 M	73 G	47
YOLOv5n	97.2	75.1	1.7 M	4.1 G	104
LMSD-YOLO *	98.0	-	3.5 M	6.6 G	68
LFer-Net *	**98.2**	**80.4**	0.6 M	1.9 G	144
LSDet (ours)	96.3	66.2	**0.3 M**	**1.2 G**	**186.3**

Note: The experimental results of methods with * are from their original articles.

To further validate the efficiency of our proposed method, we directly used YOLOv5n to process complete original images, then compared these results with those obtained using our proposed method on the same original images. The experimental results are listed in Table 4. From the table, it can be observed that our proposed method achieves higher precision and recall compared to directly using YOLOv5n. Additionally, the detection time is shorter. As is well known, the STFT and ISTFT processes are relatively time-consuming. However, in our approach, we first filter out echo signals that do not contain ships using a threshold, so only a portion of the echoes need to undergo the STFT and ISTFT processes. As a result, the total time required is less than that needed for traditional models to detect the entire image.

Table 4. Quantitative evaluation of methods solely based on deep learning and our proposed method.

Method	P	R	Times
Complete YOLOv5n	98.2	95.6	8.32 s
Ours method	100	98.7	5.47 s

Figure 10 displays the visual detection results of both methods and shows that our proposed method has fewer missed detections and false alarms.
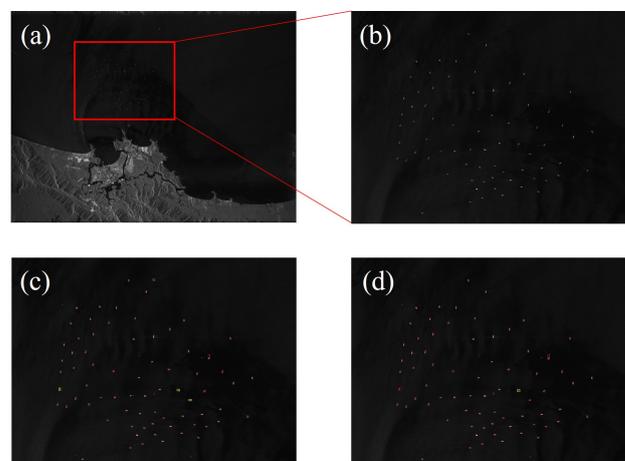
Figure 10. Visualization of the detection results. (**a**) The original image. (**b**) ROI in (**a**). (**c**) Results of direct detection using YOLOv5n on complete original images. (**d**) Detection results of our proposed method. Red: detection results. Yellow: missed detections. Blue: false alarms.

4. Conclusions

In this article, we propose a lightweight SAR ship detection method combining complex information and deep learning. To effectively leverage the complex information of SAR imagery and deep learning to address the detection difficulties caused by the characteristics of SAR imagery in the image domain, we use ESP technology to identify the presence of ships in each azimuth cell of SLC SAR images. Subsequently, we transform the azimuth cell containing the ship to the time-frequency domain by using complex information of SLC SAR imagery. The amplitude information of SLC SAR imagery in the time-frequency domain exhibits distinctive characteristics conducive to easy detection. Consequently, we constructed a small dataset named TFSID utilizing amplitude images. Exploiting its advantageous features, we designed a lightweight ship detector named LSDet, which has only about 10% as many parameters as the lightest mainstream detector. To validate the efficiency and effectiveness of our proposed method, we conducted extensive experiments on different datasets. The experimental results provide ample evidence for the feasibility and correctness of our proposed method.

Author Contributions: Conceptualization, Z.L.; Methodology, C.T., Z.L., F.X., X.W. and D.L.; Software, C.T.; Validation, C.T.; Writing—original draft, C.T.; Writing—review & editing, Z.L., F.X., X.W. and D.L.; Funding acquisition, F.X., X.W. and D.L. All authors have read and agreed to the published version of the manuscript.

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

References

1. Moreira, A.; Prats-Iraola, P.; Younis, M.; Krieger, G.; Hajnsek, I.; Papathanassiou, K.P. A tutorial on synthetic aperture radar. *IEEE Geosci. Remote Sens. Mag.* **2013**, *1*, 6–43. [CrossRef]
2. Li, J.; Xu, C.; Su, H.; Gao, L.; Wang, T. Deep learning for SAR ship detection: Past, present and future. *Remote Sens.* **2022**, *14*, 2712. [CrossRef]
3. Liu, T.; Yang, Z.; Gao, G.; Marino, A.; Chen, S.W. Simultaneous diagonalization of Hermitian matrices and its application in PolSAR ship detection. *IEEE Tran. Geosci. Remote Sens.* **2023**, *61*, 5220818. [CrossRef]
4. Yang, Z.; Fang, L.; Shen, B.; Liu, T. PolSAR Ship Detection Based on Azimuth Sublook Polarimetric Covariance Matrix. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2022**, *15*, 8506–8518. [CrossRef]
5. Goldstein, G. False-alarm regulation in log-normal and Weibull clutter. *IEEE Trans. Aerosp. Electron. Syst.* **1973**, *AES-9*, 84–92. [CrossRef]
6. Crisp, D.J. *The State-of-the-Art in Ship Detection in Synthetic Aperture Radar Imagery*; Citeseer: Princeton, NJ, USA, 2004
7. Kuttikkad, S.; Chellappa, R. Non-Gaussian CFAR techniques for target detection in high resolution SAR images. In Proceedings of the 1st International Conference on Image Processing, Austin, TX, USA, 13–16 November 1994; Volume 1, pp. 910–914.
8. Qin, X.; Zhou, S.; Zou, H.; Gao, G. A CFAR detection algorithm for generalized gamma distributed background in high-resolution SAR images. *IEEE Geosci. Remote Sens. Lett.* **2012**, *10*, 806–810.
9. Tao, D.; Anfinsen, S.N.; Brekke, C. Robust CFAR detector based on truncated statistics in multiple-target situations. *IEEE Trans. Geosci. Remote Sens.* **2015**, *54*, 117–134. [CrossRef]
10. Ai, J.; Yang, X.; Song, J.; Dong, Z.; Jia, L.; Zhou, F. An adaptively truncated clutter-statistics-based two-parameter CFAR detector in SAR imagery. *IEEE J. Ocean. Eng.* **2017**, *43*, 267–279. [CrossRef]
11. Shao, Z.; Zhang, X.; Xu, X.; Zeng, T.; Zhang, T.; Shi, J. CFAR-guided Convolution Neural Network for Large Scale Scene SAR Ship Detection. In Proceedings of the 2023 IEEE Radar Conference (RadarConf23), San Antonio, TX, USA, 1–5 May 2023; pp. 1–5.
12. Zeng, T.; Zhang, T.; Shao, Z.; Xu, X.; Zhang, W.; Shi, J.; Jun, W.; Zhang, X. CFAR-DP-FW: A CFAR-guided Dual-Polarization Fusion Framework for Large Scene SAR Ship Detection. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* **2024**, *17*, 7242–7259. [CrossRef]
13. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the 2014 IEEE International Conference on Computer Vision (ICCV), Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
14. Girshick, R. Fast r-cnn. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1440–1448.

15. 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]

16. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 11–14 October 2016; Springer: Cham, Switzerland, 2016; pp. 21–37.

17. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.

18. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.

19. Law, H.; Deng, J. Cornernet: Detecting objects as paired keypoints. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 734–750.

20. Zhou, X.; Wang, D.; Krähenbühl, P. Objects as points. *arXiv* **2019**, arXiv:1904.07850.

21. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.

22. Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; Zagoruyko, S. End-to-end object detection with transformers. In Proceedings of the European Conference on Computer Vision (ECCV), Glasgow, UK, 23–28 August 2020; Springer: Cham, Switzerland, 2020; pp. 213–229.

23. Wang, Z.; Du, L.; Mao, J.; Liu, B.; Yang, D. SAR target detection based on SSD with data augmentation and transfer learning. *IEEE Geosci. Remote Sens. Lett.* **2018**, *16*, 150–154. [CrossRef]

24. Yang, R.; Pan, Z.; Jia, X.; Zhang, L.; Deng, Y. A novel CNN-based detector for ship detection based on rotatable bounding box in SAR images. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2021**, *14*, 1938–1958. [CrossRef]

25. Sun, Z.; Leng, X.; Lei, Y.; Xiong, B.; Ji, K.; Kuang, G. BiFA-YOLO: A novel YOLO-based method for arbitrary-oriented ship detection in high-resolution SAR images. *Remote Sens.* **2021**, *13*, 4209. [CrossRef]

26. Yang, Y.; Ju, Y.; Zhou, Z. A super lightweight and efficient SAR image ship detector. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 4006805. [CrossRef]

27. Tang, X.; Zhang, J.; Xia, Y.; Xiao, H. DBW-YOLO: A High-Precision SAR Ship Detection Method for Complex Environments. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 7029–7039. [CrossRef]

28. Zhou, S.; Zhang, M.; Wu, L.; Yu, D.; Li, J.; Fan, F.; Zhang, L.; Liu, Y. Lightweight SAR Ship Detection Network Based on Transformer and Feature Enhancement. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 4845–4858. [CrossRef]

29. Zhou, Y.; Liu, H.; Ma, F.; Pan, Z.; Zhang, F. A sidelobe-aware small ship detection network for synthetic aperture radar imagery. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5205516. [CrossRef]

30. Liu, Y.; Yan, G.; Ma, F.; Zhou, Y.; Zhang, F. SAR Ship Detection Based on Explainable Evidence Learning under Intra-class Imbalance. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5207715.

31. Wu, W.; Li, X.; Guo, H.; Ferro-Famil, L.; Zhang, L. Noncircularity parameters and their potential applications in UHR MMW SAR data sets. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1547–1551. [CrossRef]

32. El-Darymli, K.; Mcguire, P.; Gill, E.W.; Power, D.; Moloney, C. Characterization and statistical modeling of phase in single-channel synthetic aperture radar imagery. *IEEE Trans. Aerosp. Electron. Syst.* **2015**, *51*, 2071–2092. [CrossRef]

33. Leng, X.; Ji, K.; Zhou, S.; Xing, X. Ship detection based on complex signal kurtosis in single-channel SAR imagery. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 6447–6461. [CrossRef]

34. Lv, Z.; Lu, J.; Wang, Q.; Guo, Z.; Li, N. ESP-LRSMD: A Two-Step Detector for Ship Detection Using SLC SAR Imagery. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5233516. [CrossRef]

35. Rao, B.D.; Hari, K. Weighted subspace methods and spatial smoothing: Analysis and comparison. *IEEE Trans. Signal Process.* **1993**, *41*, 788–803. [CrossRef]

36. Zhuang, X.; Cui, X.; Lu, M.; Feng, Z. Low-complexity method for DOA estimation based on ESPRIT. *J. Syst. Eng. Electron.* **2010**, *21*, 729–733. [CrossRef]

37. Gabor, D. Theory of communication. Part 1: The analysis of information. *J. IEEE* **1946**, *93*, 429–441. [CrossRef]

38. Sunkara, R.; Luo, T. No more strided convolutions or pooling: A new CNN building block for low-resolution images and small objects. In Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Grenoble, France, 19–23 September 2022; Springer: Cham, Switzerland, 2022; pp. 443–459.

39. Zhang, R. Making convolutional networks shift-invariant again. In Proceedings of the 36th International Conference on Machine Learning (ICML), Long Beach, CA, USA, 9–15 June 2019; PMLR: Birmingham, UK, 2019; pp. 7324–7334.

40. Yu, W.; Zhou, P.; Yan, S.; Wang, X. Inceptionnext: When inception meets convnext. *arXiv* **2023**, arXiv:2303.16900.

41. Zhang, T.; Zhang, X.; Li, J.; Xu, X.; Wang, B.; Zhan, X.; Xu, Y.; Ke, X.; Zeng, T.; Su, H.; et al. SAR ship detection dataset (SSDD): Official release and comprehensive data analysis. *Remote Sens.* **2021**, *13*, 3690. [CrossRef]

42. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021; pp. 10012–10022.

43. Tian, Z.; Shen, C.; Chen, H.; He, T. FCOS: Fully Convolutional One-Stage Object Detection. *arXiv* **2019**, arXiv:1904.01355.

44. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2961–2969.

45. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, 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. 2980–2988.
46. 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 theIEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 7464–7475.
47. Guo, Y.; Chen, S.; Zhan, R.; Wang, W.; Zhang, J. LMSD-YOLO: A lightweight YOLO algorithm for multi-scale SAR ship detection. *Remote. Sens.* **2022**, *14*, 4801. [CrossRef]
48. Tian, C.; Liu, D.; Xue, F.; Lv, Z.; Wu, X. Faster and Lighter: A Novel Ship Detector for SAR Images. *IEEE Geosci. Remote Sens. Lett.* **2024**, *21*, 4002005. [CrossRef]