Article

Ship Formation Identification with Spatial Features and Deep Learning for HFSWR

Jiaqi Wang , **Aijun Liu** * , **Changjun Yu and Yuanzheng Ji**

School of Information Science and Engineering, Harbin Institute of Technology at Weihai, Weihai 264209, China; 21b905018@stu.hit.edu.cn (J.W.); yuchangjun@hit.edu.cn (C.Y.); 22s130417@stu.hit.edu.cn (Y.J.)
* Correspondence: liuaijun@hit.edu.cn

Abstract: Ship detection has been an area of focus for high-frequency surface wave radar (HFSWR). The detection and identification of ship formation have proven significant in early warning, while studies on the formation identification are limited due to the complex background and low resolution of HFSWR. In this paper, we first establish a spatial distribution model of ship formation in HFSWR. Then, we propose a cascade identification algorithm of ship formation in the clutter edge. The proposed algorithm includes a preprocessing stage and a two-stage formation identification stage. The Faster R-CNN is introduced in the preprocessing stage to locate the clutter regions. In the first stage, we propose an extremum detector based on connected regions to extract suspicious regions. The suspicious regions contain ship formations, single-ship targets, and false targets. In the second stage, we design a network connected by a convolutional neural network (CNN) and an extreme learning machine (ELM) to identify two densely distributed ship formations from inhomogeneous clutter and single-ship targets. The experimental results based on the factual HFSWR background demonstrate that the proposed cascade identification algorithm is superior to the extremum detector combined with the classical CNN algorithm for ship formation identification. Meanwhile, the proposed algorithm performs well in weak formation and deformed formation identification.

Keywords: high-frequency surface wave radar (HFSWR); cascade identification algorithm; ship formation identification; deep learning

1. Introduction

Because of its beyond-the line-of-sight monitoring and all-weather-continuous remote sensing [1,2], high-frequency surface wave radar (HFSWR) has gained considerable attention. HFSWR has been widely used for two dominant areas: real-time maritime surveillance [3] and sea surface target detection [4,5]. Among these, ship detection has been a focus of scholars and researchers [6,7]. The Range-Doppler (RD) image background contains various clutter mainly composed of sea clutter, ionospheric clutter, and dense targets' interference [8]. The existing research mainly focuses on continuously improving the target resolution ability.

The constant false alarm rate (CFAR) is a valuable point target detection technique based on comparing the neighboring reference cells' mean energy [9,10]. This technique has been used in ship extraction by Hinz [11] and Liu et al. [12]. Dzvonkovskaya et al. [13] analyzed the characteristics of sea clutter, meteor clutter, noise, and land reflection clutter and proposed an adaptive thresholding estimation strategy for HFSWR ship detection. In the case of two targets in the same cell, Rohling H [14] proposed ordered statistical CFAR (OS-CFAR), which could effectively reduce the impact on the detection probability and the false alarm probability caused by inhomogeneous spatial distribution in local regions. Li et al. [15,16] improved the CFAR detector using spatial correlation as prior information. Experiments based on the CFAR detector have shown that spatial information is beneficial for improving the accuracy of dense targets detection.

Jangal et al. [17] transformed the detection based on the spatial distribution characteristics of targets and clutter into image processing based on the morphology and distribution characteristics. They introduced the idea of wavelet-transform and proposed a target signal extraction method based on image decomposition. Researchers have combined the image processing technique with RD images of HFSWR and continuously improved them, achieving good results in target detection and clutter suppression. Baussard et al. [18] analyzed the morphological features of the measured RD images and used curve analysis methods to remove components that interfere with target detection in the images. Subsequently, a ship detection technique based on the authentic RD images of HFSWR was proposed by Baussard et al. [19], which exploits multi-scale variation and sparse representation morphological component analysis and obtains correct target detection results. Li et al. [20] proposed a target detection method based on discrete wavelet transform (DWT), which automatically determines the optimal scale of wavelet transform instead of selecting based on experience.

Ocean surveillance using HFSWR is commonly performed in wide-range RD images. The targets to be detected in the RD images are very tiny. Regarding target detection, simple image processing methods rely on manual feature extraction and need more robustness. According to the existing HFSWR research, a two-stage intelligent processing algorithm is very suitable for the automatic extraction of an RD image's optimal spatial features and performs well [21–23]. In the first stage, many targets are quickly located, ensuring a high detection rate. In the second stage, false targets are removed. Existing research assumes that ships are independently distributed and treated as point targets. The information obtained in the study of isolated targets is limited. In HFSWR, the echoes from a single large-sized ship are concentrated in a single cell of the RD images and can be considered stable in a short period [24]. Echo signals of ships sailing in a dense formation or group typically occupy multiple adjacent cells and present a specific spatial geometric shape. Peng et al. [25] established a narrowband coherent radar multi-aircraft formation (MAF) echo model. They applied polynomial Fourier transform (PFT) to accurately identify the number of aircraft and the motion parameters of each aircraft. The experiments in a clean background have proven the effectiveness of the proposed method. Liang et al. [26] simulated the over-the-horizon radar (OTHR) echo images of MAF with multipath effect using spectral color blocks number and amplitude. They adopted a convolutional neural network (CNN) to recognize the number of aircraft and conducted experiments in homogeneous clutter. In the complex background of HFSWR, research on formation detection and identification is limited.

In this paper, we propose a novel cascade identification algorithm for ship formation in the clutter edge. We first analyze the motion of HFSWR ship formation and establish a matching ship formation spatial distribution model. There are two types of formations with rigid structures that appear as extended targets with specific spatial shapes on the RD images. Then, we introduce the Faster R-CNN to locate the clutter regions and propose a two-stage formation identification algorithm. In the first stage, the extremum detector based on the gray value is employed to find cells with suspicious targets, and the Seed-Filling (SF) algorithm is introduced to connect multi-cell regions belonging to the same ship formation. An extremum detector based on connected regions is proposed to avoid duplicate detection of the same ship formation and reduce the data to be processed in the second stage. In the second stage, a lightweight CNN is designed to extract the convolutional features of suspicious targets. The highly abstract features extracted by CNN are input into the extreme learning machine (ELM) for efficient classification. We propose CNN-ELM to identify two densely distributed ship formations from inhomogeneous clutter and single-ship targets. The experimental results based on the factual HFSWR background demonstrate that our CNN-ELM model surpasses the classic Alexnet and Resnet18 in terms of the detection accuracy (97.5%) and processing time (0.871 s). Our proposed cascade identification algorithm performs excellently on datasets consisting of two types of ship formations and single-ship targets. Meanwhile, the proposed algorithm is tested on weak ship formation datasets (Weak set-1 and Weak set-2) and deformed formation datasets

(Deformed set-1 and Deformed set-2). The proposed algorithm achieves an identification accuracy of 82% with an average signal-to-clutter ratio (SCR) of 5–7.5 dB. The proposed algorithm achieves an identification accuracy of 73.38% when the sub-targets deviate by 20% and an identification accuracy of 96.25% through transfer learning. Compared with the extremum detector combined with the classical CNN algorithm, our proposed algorithm performs excellently.

The rest of the paper is organized as follows. In Section 2, we describe the spatial distribution model of ship formation. In Section 3, we describe the proposed cascade identification algorithm. In Section 4, we provide the experimental results and discussion. In Section 5, we conclude.

2. Spatial Distribution Model of Ship Formation

Figure 1 shows the two-dimensional geometric model of several ship formations with multi-ships. The model discussed in this paper is universal and not limited to the following types. The ship formation movement observed by HFSWR can be analyzed from two perspectives: the formation center movement and sub-targets' movement. The formation center is consistent with the overall ship formation movement, which has an absolute motion state. Sub-targets' motion can be decomposed into absolute motion consistent with the ship formation motion, and relative motion relative to the formation center.
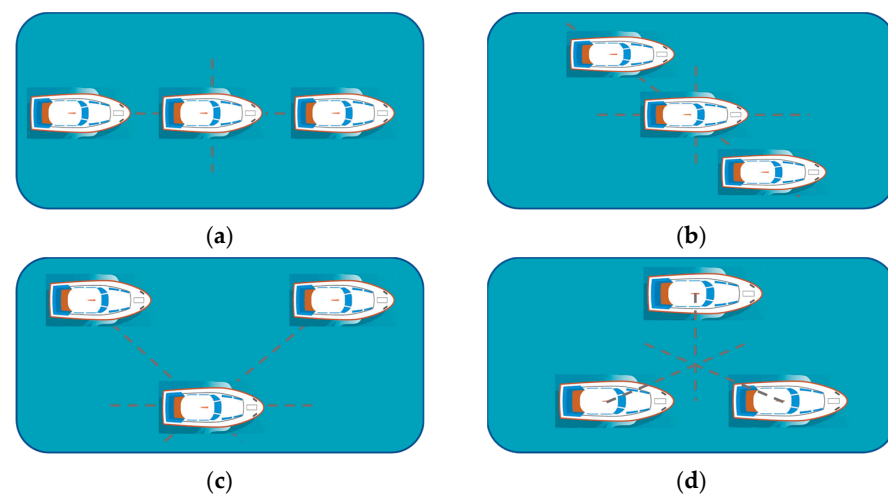

Figure 1. Two-dimensional geometric model of several ship formations: (**a**) Single-column formation; (**b**) bias-column formation; (**c**) V-shaped formation; (**d**) centrosymmetric formation.

Referring to the inverse synthetic aperture radar (ISAR) imaging model [27,28], the relative motion of sub-targets can be abstracted as a rotational motion around the formation center. Figure 2 shows the motion model of the ship formation. The HFSWR is located at the origin O of the global coordinate system XOY. The X axis is tangent to the coast, and the Y axis is perpendicular to the X axis. The ships in formation are simplified as point targets, with one sub-target located at p_1. The formation center o' is the reference point for measuring the position relationship of all sub-targets. The local coordinate system $xo'y$ matches the distribution of ship formation, which is related to the control system of the formation. There is counterclockwise rotation at an angle of α from the global coordinate system XOY to the local coordinate system $xo'y$. The coordinate of sub-target p_1 in the local coordinate system $xo'y$ is (x,y). The radar line of sight (RLOS) coordinate system $ro'l$ takes o' as the coordinate origin, where the r axis is consistent with the RLOS direction. The azimuth angle of the formation center o' relative to the HFSWR varies between $0°$ and $180°$. The ship formation with a constant velocity v_{ship} sails along the X axis direction of the global coordinate system. Their slowly changing heading can be assumed to be stable during the coherent integration time (CIT).
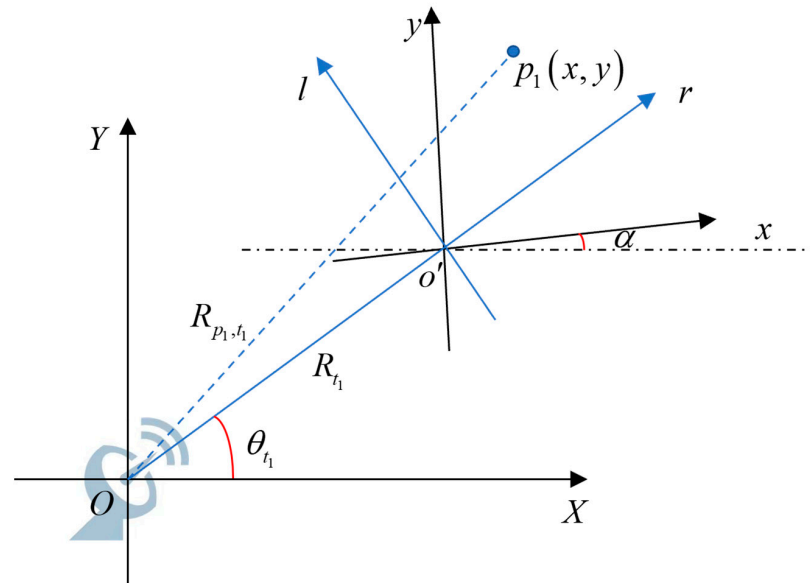
Figure 2. Motion model of the ship formation.

At time t_1, the distance from the sub-target p_1 to the HFSWR can be expressed as:

$$R_{p_1,t_1} = \|\mathbf{A}\|, \tag{1}$$

where

$$\begin{aligned}
\mathbf{A} &= R_{t_1}\mathbf{V} + \mathbf{U}\begin{bmatrix} x \\ y \end{bmatrix} \\
\mathbf{V} &= [\cos\theta_{t_1}, \sin\theta_{t_1}]^T \\
\mathbf{U} &= [\cos\alpha, -\sin\alpha; \sin\alpha, \cos\alpha]^T
\end{aligned}, \tag{2}$$

R_{t_1} is the distance between the formation center o' and the radar at time t_1, and θ_{t_1} is the azimuth angle of the o' relative to the radar at time t_1. During the beyond-the line-of-sight monitoring of HFSWR, the absolute coordinate value of p_1 is much smaller than R_{t_1}. The R_{p_1,t_1} can be approximated as:

$$\begin{aligned}
R_{p_1,t_1} &\approx R_{t_1} + \mathbf{W}\begin{bmatrix} x \\ y \end{bmatrix} \\
\mathbf{W} &= [\cos(\theta_{t_1} - \alpha), \sin(\theta_{t_1} - \alpha)]
\end{aligned}. \tag{3}$$

At time $t_2 = t_1 + \triangle t$, where $\triangle t$ is the time interval, the azimuth angle of o' relative to the radar can be expressed as:

$$\theta_{t_2} = \theta_{t_1} + \triangle\theta, \tag{4}$$

where $\triangle\theta = \arcsin(b_1/\sqrt{b_1 + b_2})$ is the azimuth change during the time $\triangle t$, $b_1 = R_{t_1}\sin\theta_{t_1}$, and $b_2 = R_{t_1}\cos\theta_{t_1} + v_{ship}\triangle t$. Usually, $\triangle t$ and $\triangle\theta$ are very small. At time t_2, the distance between p_1 and the radar can be approximated as:

$$R_{p_1,t_2} \approx R_{t_2} + r_{t_1} - l_{t_1} \cdot \triangle\theta, \tag{5}$$

where (r_{t_1}, l_{t_1}) is the coordinate of p_1 in the $ro'l$ coordinate system, and it can be expressed as:

$$\begin{aligned}
\begin{bmatrix} r_{t_1} \\ l_{t_1} \end{bmatrix} &= \mathbf{H}\begin{bmatrix} x \\ y \end{bmatrix} \\
\mathbf{H} &= \begin{bmatrix} \cos(\theta_{t_1} - \alpha) & \sin(\theta_{t_1} - \alpha) \\ -\sin(\theta_{t_1} - \alpha) & \cos(\theta_{t_1} - \alpha) \end{bmatrix}
\end{aligned}, \tag{6}$$

$R_{t_2} = R_{t_1} + v_{r,t_1} \triangle t$ is the distance between o' and the radar at time t_2, and $v_{r,t_1} = v_{ship} \cos \theta_{t_1}$ is the radial velocity of p_1 relative to the radar.

The distance between the sub-targets and the HFSWR can be obtained through iteration during the CIT. The transmission waveform of HFSWR is a frequency-modulated interrupted continuous wave (FMICW). Assuming $t_1 = jT_c$ is the end time of the jth sweep, where T_c is the sweep width, the $(j + 1)$th sweep time can be represented by

$$t_2 = t_1 + kq + \frac{\hat{t}}{T_0}, \tag{7}$$

where q is the pulse period, and T_0 is the pulse width in seconds. The transmission waveform can be expressed as:

$$s_{t_2} = A \exp\left(j2\pi\left(f_c t_2 + \frac{B}{2T_c} t_2^2\right)\right), \tag{8}$$

where A is the amplitude of transmission signal, f_c is the carrier frequency, and B is the bandwidth. The mixing output signal of the sub-target p_1 can be expressed as:

$$s_{p_1} = C \exp(j2\pi\varphi_{p_1}), \tag{9}$$

where C is the amplitude of the echo signal. The phase of the echo signal can be expressed as:

$$\varphi_{p_1} = f_1(R_{t_1}, R_{t_2}) + f_2(r_{t_1}, l_{t_1}), \tag{10}$$

where the first item

$$f_1(R_{t_1}, R_{t_2}) = \frac{2BR_{t_1}}{cT_c} t_2 + \frac{2f_c R_{t_2}}{c}, \tag{11}$$

is related to the distance between the formation center and the radar, and the second item

$$f_2(r_{t_1}, l_{t_1}) = \frac{2Br_{t_1}}{cT_c} t_2 + \frac{2f_c r_{t_1} - \triangle\theta l_{t_1}}{c}, \tag{12}$$

is related to the coordinate of p_1 in the $ro'l$ coordinate system. The echo signal of the ship formation can be expressed as:

$$s_{IF} = C \sum_{i=1}^{N} \exp(j2\pi\varphi_{p_i}), \tag{13}$$

where N is the number of sub-targets in the formations. The main influencing factors of amplitude C include the transmission power, radar cross section (RCS), and distance. The echo amplitude of sub-targets is considered to be the same constant. The signal model is related to sub-target distribution and the ship formation motion.

During the CIT, the ship formation spatial distribution model is obtained by imaging the echo signal of multiple sweeps.

3. Proposed Cascade Identification Algorithm

In this section, we introduce the proposed cascade identification algorithm, which detects and identifies ship formations sparsely injected into the real clutter edges.

3.1. Cascade Identification Algorithm Framework and Preprocessing

Figure 3 shows the framework of the proposed algorithm. In the preprocessing stage, the Faster R-CNN is introduced to locate the clutter region of the collected RD images. In the first stage, an extremum detector based on connected regions is proposed to extract targets. The processing principle in this stage is to avoid losing targets, quickly obtain all suspicious targets, and group the ship formation for output. In the second stage, a CNN-ELM is designed to eliminate false targets and identify two densely distributed ship formations.
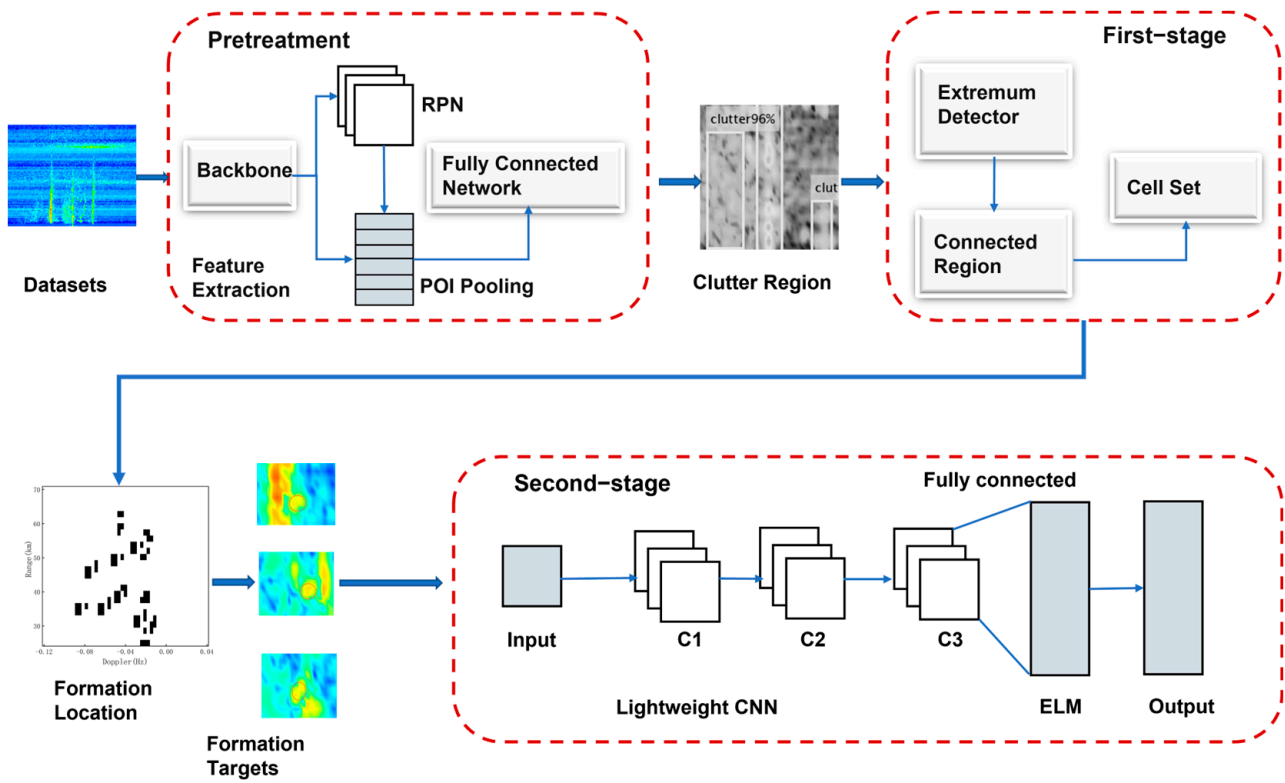
Figure 3. Framework of the proposed cascade identification method.

Compared to color images, grayscale images are less space-consuming and retain essential information, which is suitable for real-time processing. Before the preprocessing stage, the color RD images are transformed into grayscale images, and the grayscale value of each pixel can be expressed as

$$g_x(i,j) = W_R R_x(i,j) + W_G G_x(i,j) + W_B B_x(i,j), \tag{14}$$

where W_R, W_G, and W_B are the weights of R, G, and B, respectively.

In the preprocessing stage, Faster R-CNN is merely used to distinguish the clutter regions and locate them. Faster R-CNN is a target classifier in multi-region-based proposals [29], and attention-oriented Region Proposal Networks (RPN) extract the interest regions. Using convolutional image feature maps shared with RPN, the fully connected layer can adopt a deep network and improve the quality of proposals.

The default feature extraction and target classification network of the Faster R-CNN is the VGG-16. The VGG-16 consists of 5 convolutional layers and 13 shared layers, suitable for large-volume datasets with over 10^4 images. The dataset composed of images mentioned in Section 4.2 cannot meet the training requirements of VGG-16. We replaced the original VGG-16 with Resnet50. The Resnet50 introduces residual modules to avoid the problems of gradient vanishing and model degradation.

3.2. Extremum Detector Based on Connected Region

The grayscale values of target and sea clutter are lower than in the background. Ship formation occupies more stable high-energy cells than the single-ship target, which appears as an isolated point. In this stage, an extremum detector based on connected regions is adopted.

First, the discrimination criteria for the extremum detector are expressed as:

$$g(x_{test}) = \begin{cases} 1, & h(x_{test}) \leq \lambda k_0 \\ 0, & h(x_{test}) > \lambda k_0 \end{cases}, \tag{15}$$

where λ is the threshold factor, k_0 is the average grayscale value of the reference cells, and $h(x_{test})$ is the grayscale value of the tested cell x_{test}. 1 indicates the cell is a suspicious target, and 0 indicates the background. We adjust the threshold factor λ to ensure the extremum detector detects all targets.

Then, the SF algorithm is introduced to connect the cells with suspicious targets. The cells with $g(x_{test}) = 1$ are seeds, of which the grayscale value is set to 0. We traverse the four adjacent cells around the seeds, including the top, bottom, left, and right, and merge the cells with the grayscale value of the 0 value into a connected region. We search for connected regions using the newly merged cells as seeds. The detection process will continue until all the cells with suspicious targets have been traversed.

3.3. Lightweight CNN-ELM

CNN performs well in extracting image features and identifying real targets contaminated with clutter [21]. ELM is a feedforward network with a single hidden layer, which has a simple and effective training method and can quickly obtain local optimal solutions. We design a CNN-ELM classifier to identify ship formations in the clutter edge. The structure of the lightweight CNN is based on Alexnet and was determined after extensive debugging. To accelerate the network speed, the designed CNN is shallower than Alexnet. The lightweight CNN consists of three convolution layers, two pooling layers, and two fully connected layers. A dropout layer was added to the third convolution layer to prevent overfitting. The network structure is given in Table 1. Compared to Alexnet, the proposed lightweight CNN has fewer pre-training parameters.

Table 1. Whole CNN structure of the lightweight CNN.

| Layer	Layer Name	Parameter
1	Convolution	$32 \times 3 \times 3$ convolutions, stride [1 1]
2	ReLu	ReLu
3	Convolution	$32 \times 3 \times 3$ convolutions, stride [1 1]
4	ReLu	ReLu
5	Maxpooling	2×2 max pooling, stride [2 2]
6	Convolution	$64 \times 3 \times 3$ convolutions, stride [1 1]
7	ReLu	ReLu
8	Maxpooling	2×2 max pooling, stride [2 2]
9	DropoutLayer	0.5 dropout probability
10	Fully Connected	64 fully connected layers
11	ReLu	ReLu
12	Fully Connected	128 fully connected layers

4. Experiment and Discussion

In order to demonstrate the effectiveness of the proposed cascade identification algorithm, experiments are conducted using simulated targets and measured RD image backgrounds. We present the experimental parameters including the measured background and different types of formations, in Section 4.1. The RD data structure and multiple datasets are introduced in Section 4.2. The model training and formation identification are presented in Section 4.3. We also evaluate the performance of the proposed algorithm in the case of weak and deformed formation identification in Section 4.4.

4.1. Experiment Parameters

The origin RD data, composed of various clutter and targets, are obtained through the HFSWR system located in the city of Weihai. The HFSWR system parameters are given in Table 2. The real targets are artificially removed from the measured data, and simulated targets, including two types of ship formations shown in Figure 4 and single-ship targets, are injected into the clutter edge position of RD images.

Table 2. HFSWR system parameters.

Parameter Name	Parameter Symbol	Parameter Value
Carrier frequency	f_c	4.5 MHz
Bandwidth	B	40 KHz
Pulse width	T_0	0.35 ms
Sampling frequency	f_0	24 MHz
Range resolution	/	1.88 km
Doppler resolution	/	0.0039 hz
CIT	T_{total}	249.98 s
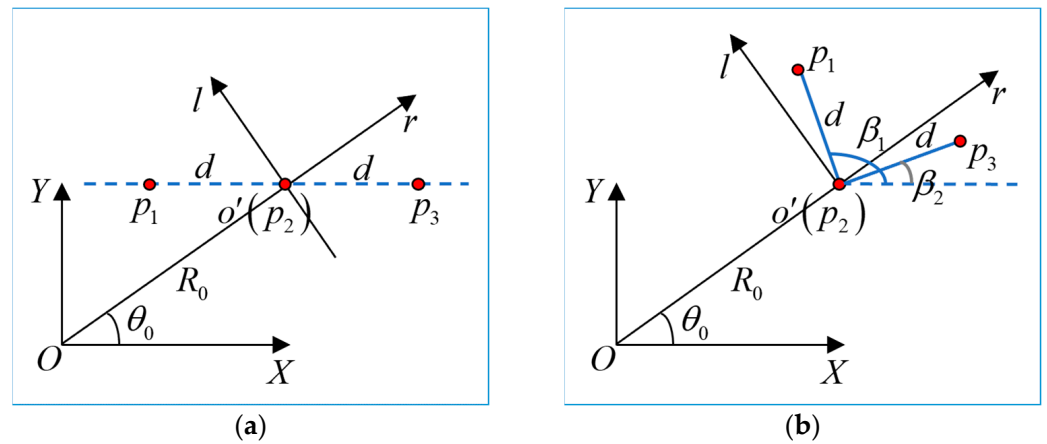
Figure 4. Geometric models of two ship formations: (**a**) geometric models of single-column formation; (**b**) geometric models of V-shaped formation.

The geometric parameters of the single-column formation and the V-shaped formation are shown in Table 3. Figure 5 shows the spatial distribution model of the two formations without clutter. Figure 6 shows the distribution of the single-column and V-shaped formation in the RD images. The constructed RD images cover widely in the range and Doppler domains. The images have been cropped for ease of display in Figures 5 and 6. The two types of ship formations have the same parameters, except for the formation shape. For the single-column formation, the maximum difference between the sub-targets and the formation center in the range and Doppler domains is 1.03 km and 0.0022 Hz, respectively. For the V-shaped formation, the maximum distance of the sub-targets and formation center in the range domain is 1.75 km, and the gap between p_1 and p_3 in the Doppler domain is 0.0019 Hz. There is less difference between sub-targets in the range domain and Doppler domain than in the HFSWR resolution given in Table 2. Formations are presented as extended targets with slight differences in outline.

Table 3. Simulation parameters of ship formations.

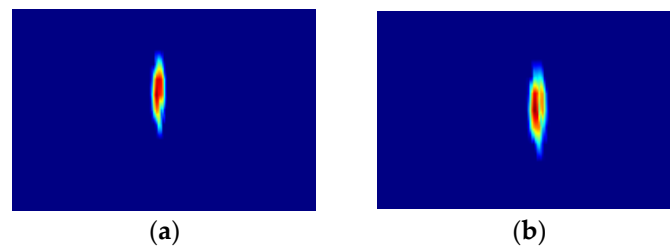| Parameter Name	Parameter Symbol	Parameter Value
Initial distance	R_0	100 km
Initial azimuth angle	θ_0	60°
Sailing speed	v_{ship}	8 m/s
Distance between formation center and sub-target	d	2 km
Deflection angle	β_1	120°
Deflection angle	β_2	30°

Figure 5. Spatial distribution model of ship formations without clutter: (**a**) spatial distribution model of single-column formation; (**b**) spatial distribution model of V-shaped formation.
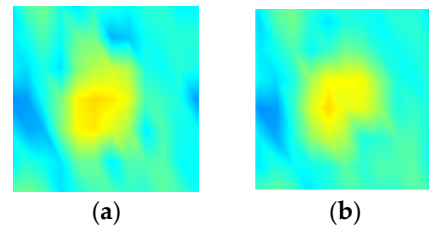

Figure 6. Spatial distribution model of ship formations in clutter edge: (**a**) spatial distribution model of single-column formation; (**b**) spatial distribution model of V-shaped formation.

We also analyze the distribution feature of the weak formation and deformed formation. The target energy is measured by the SCR, and the SCR is expressed as:

$$SCR \;=\; 10 \cdot \log_{10}\left(\frac{P_s}{P_{ac}}\right), \tag{16}$$

where P_s is the amplitude of the target center, and P_{ac} is the average amplitude of the reference clutter. Figure 7 shows ship formations with low SCR. Moreover, the ship formations with SCR of 5 and 7.5 dB almost blend into the background.
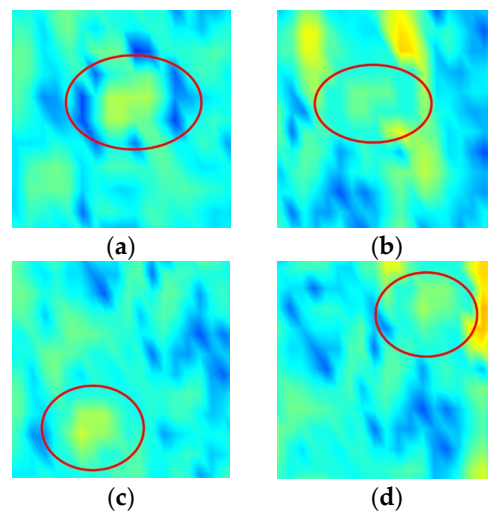

Figure 7. Spatial distribution model of ship formations with low SCR: (**a**) single−column formation with SCR of 7.5 to 10 dB; (**b**) single−column formation with SCR of 5 to 7.5 dB; (**c**) V−shaped formation with SCR of 7.5 to 10 dB; (**d**) V−shaped formation with SCR of 5 to 7.5 dB.

Figure 8 shows the deformed formation. Based on the original values, the distance between the formation center and the sub-targets is increased by 5% and 20%. Ship formations change very little at the deviation of 5%. At the deviation of 20%, the single-column formation is stretched, and the sub-targets of the V-shaped formation can be distinguished.
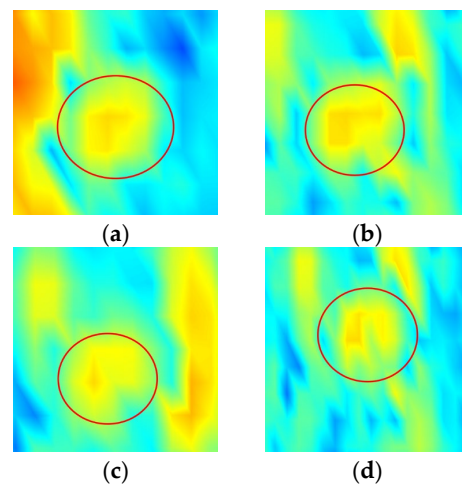
Figure 8. Spatial distribution model of ship formations under different deviations: (**a**) single−column formation at the deviation of 5%; (**b**) single−column formation at the deviation of 20%; (**c**) V−shaped formation at the deviation of 5%; (**d**) V−shaped formation at the deviation of 20%.

4.2. Data Structure and Dataset

The size of the RD images constructed in Section 4.1 is 4480×300 pixels. The color RD images are transformed into grayscale images and fed into the pre-trained Faster R-CNN to extract clutter regions. The extremum detector processes the clutter region images captured from the RD images. The first stage output consists of multiple connected regions, each represented by a set of cells from the same ship formation or a single-ship target. In the second stage, a fixed-size 2D sliding window is selected to capture target data from the images, with the center of the sliding window located at the center of the connected regions. Considering the size criteria of the window in detectors [15,16] and the characteristics of our data, the range and Doppler direction of the sliding window are set to 20 cells. The output of the CNN-ELM classifier is the identification results.

The preprocessing stage focuses on quickly extracting clutter regions rather than a profound study on the Faster R-CNN algorithm. We create a Faster R-CNN dataset containing 20 RD images and label the sea clutter. In the first stage, 20 RD images are selected as test samples, each with 45 targets at the edge of the clutter. The sample set is used to determine the threshold factor λ of the extremum detector. In the second stage, the dataset consists of equal proportional single-ship targets, single-column formations, V-shaped formations, and backgrounds. The SCR of the targets ranges from 5 dB to 25 dB. The dataset contains 4000 images and is randomly split into three parts: three-fifths of which is used as a training set, one-fifth of which is used as a validation set, and one-fifth of which is used as a testing set.

We also construct datasets for the weak and deformed formation identification separately. Each dataset contains 20 RD images, each with 45 targets at the edge of the clutter. The proportion of single-ship targets, single-column formations, and V-shaped formations is consistent. In the "Weak set-1" and "Weak set-2", the SCR of the targets is 7.5 to 10 dB and 5 to 7.5 dB, respectively. In the "Deformed set-1" and "Deformed set-2", the sub-targets are deviated by 5% and 20%, respectively.

4.3. Experimental Results

Faster R-CNN is pre-trained on ImageNet. According to the size of the clutter regions in multiple RD images, the anchors of the Faster R-CNN are adjusted to three scales (642,1282, and 2582 pixels) and three aspect ratios (0.5, 1.0, and 2.0). The anchors match the real target box with an overlap of more than 0.5. Figure 9 shows the clutter detection results of the Faster R-CNN. The bright vertical bar regions in the images are identified as sea clutter regions. The scores on the top indicate that the network has good recognition performance for sea clutter and meets the requirements of the preprocessing stage.

Figure 10 shows the grayscale distribution of single-ship target and single-column formation. The closer the cell is to the center regions, the lower the grayscale value and the higher the energy.

Ship formation occupies more stable high-energy cells than the single-ship target, which appears as an isolated point.
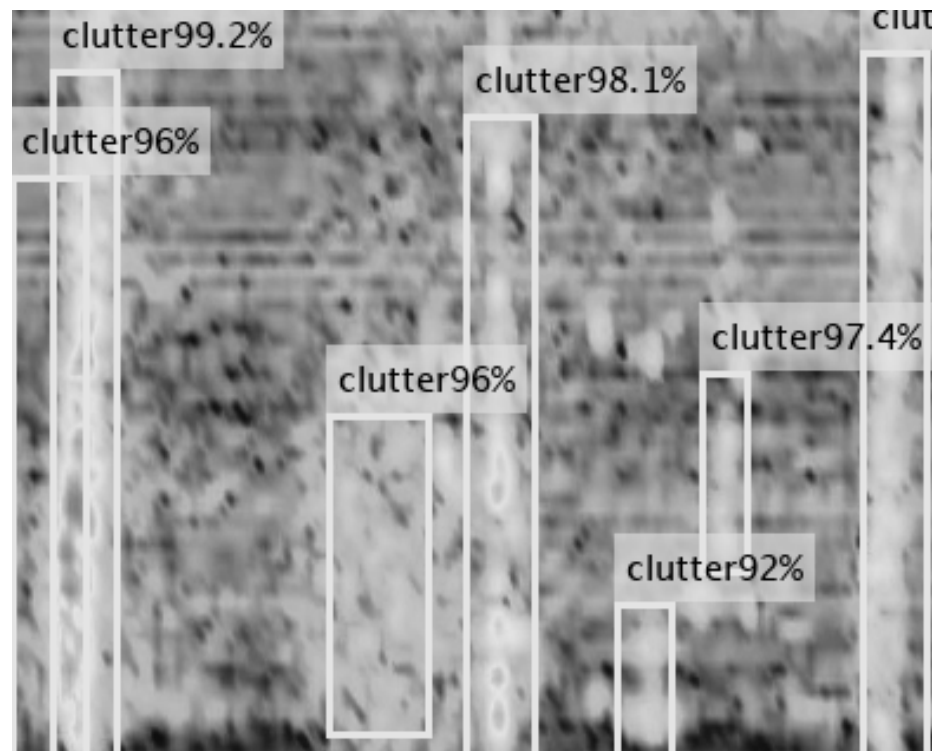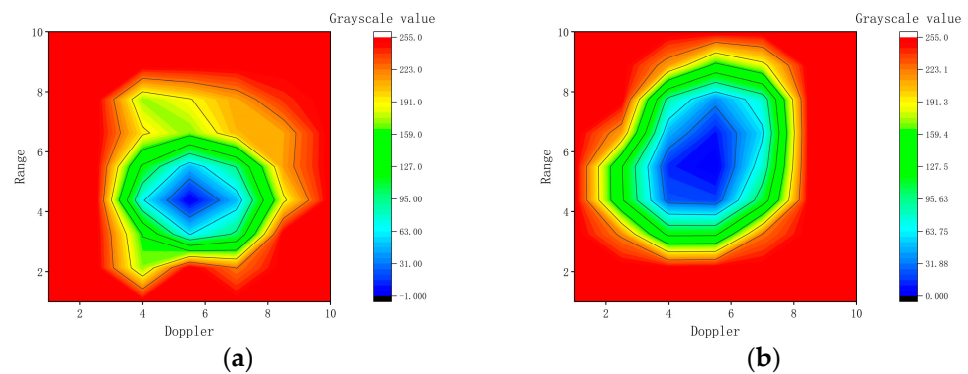

Figure 9. Detection results of the Faster R-CNN.

Figure 10. Grayscale distribution: (**a**) grayscale distribution of a single−ship target; (**b**) grayscale distribution of a ship formation.

The threshold factor λ is experimentally determined to be 0.9, and the extremum detector can detect all targets. Figure 11 shows an example of the target detection results, and the connected regions shown in black contain potential targets. From the detection results, targets are hit without any omissions. Taking Figure 11 as an example, the output of a pure extremum detector is separate cells with a quantity of 65, and the output of the connected region extremum detector is cell sets with an amount of 23. The two types of extremum detectors obtain the same cells but different output forms. The pure extremum detector extracts RD region images with 65 separate cells as the image center. In contrast, the extremum detector proposed extracts images with the cell set center as the image center. The RD region images to be classified in the second stage are reduced by 182.61% using the proposed first stage detector. From the above analysis, we have concluded that the proposed extremum detector can improve the efficiency of the cascade identification algorithm.
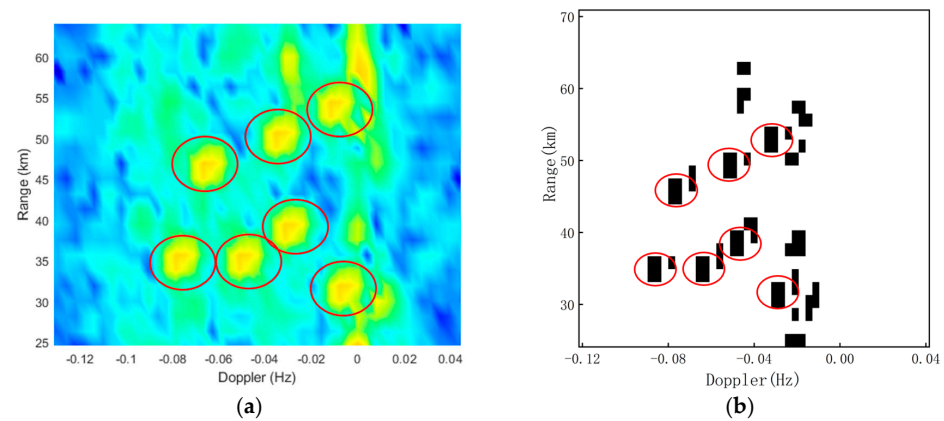
Figure 11. An example of the detection results in first stage: (**a**) RD images; (**b**) the output of the connected region extremum detector.

We train the CNN-ELM step by step, and the process is as follows:

- Train the lightweight CNN using labeled datasets, and adjust the parameters until the training accuracy reaches 80%;
- Replace the output layer of the CNN with the ELM, and train the ELM using highly abstract features extracted by the CNN until the expected results are achieved;
- Obtain a hybrid CNN-ELM model.

Based on the features of the input and empirical data, the parameters, including the initial learning rate and batch size, are adjusted to minimize training losses. The batch size of 128 and initial learning rate of 0.001 are finally applied to the CNN-ELM model, considering the training effectiveness and the running time.

The proposed CNN-ELM is compared with the classical Alexnet and the Resnet18 used in [26]. The Alexnet consists of five convolution layers and three fully connected layers, with activation functions for ReLU and Dropout layers to enhance the model generalization ability. The Resnet18 has a shallow depth and contains four stacked residual blocks. Precision, recall, accuracy, and processing time are used as the network evaluation metrics. The precision and recall can be expressed as:

$$\begin{aligned} Precision &= \frac{TP}{TP+FP} \\ Recall &= \frac{TP}{TP+FN} \end{aligned} \tag{17}$$

where TP is the number of true targets detected, TN is the number of true targets missed, and FN is the number of false targets detected. The multi-category problem is transformed into a binary classification problem, where the precision and recall of each category are calculated separately. The accuracy is expressed as the ratio of the number of correctly classified samples to the total number of samples, used to describe the global accuracy of the model. The processing time refers to the average processing time of each image, expressed as the ratio of the testing set processing time to the number of test samples. Table 4 shows the performances of the three networks.

Table 4. Performances of the three networks.

Class	Evaluation Metrics	CNN-ELM	Alexnet	Resnet18
Single-column Formation	Precision	97.01%	97.46%	95.45%
	Recall	97.50%	96.00%	94.50%
V-shaped Formation	Precision	98.02%	98.03%	96.04%
	Recall	99.00%	99.50%	97.00%
Single-ship Target	Precision	97.46%	96.50%	92.61%
	Recall	96.00%	96.50%	94.00%
/	Accuracy	97.50%	97.33%	95.17%
/	Processing Time	0.871 s	1.302 s	1.188 s

Table 4 shows that the three networks can distinguish single-ship targets, two types of dense ship formations, and backgrounds, with a classification accuracy of over 90%. From the first seven rows of Table 4, both our proposed network and Alexnet are on the level of precision and recall parameters and are superior to Resnet18. In addition, the proposed network achieves a slightly better performance in accuracy than Alexnet. The reason is that the lightweight CNN met the feature extraction requirements of the target morphology and energy, and the ELM improved the generalization ability and noise resistance. Regarding training and processing time, our proposed network is significantly better than Alexnet and Resnet18. This is due to the CNN-ELM having a shallower network structure. In summary, the designed CNN-ELM performs well.

We compare the proposed cascade identification algorithm with the extremum detector−Alexnet and extremum detector−ResNet18. The targets and identification results are annotated in the RD images. In Figure 12, the single-ship targets are marked by boxes, circles mark the single-column formations, and triangles mark the V-shaped formations. The results indicate that the proposed algorithm can correctly identify ship formations and single-ship targets contaminated with clutter. The extremum detector−Alexnet and extremum detector−Resnet18 could only identify part of the targets.
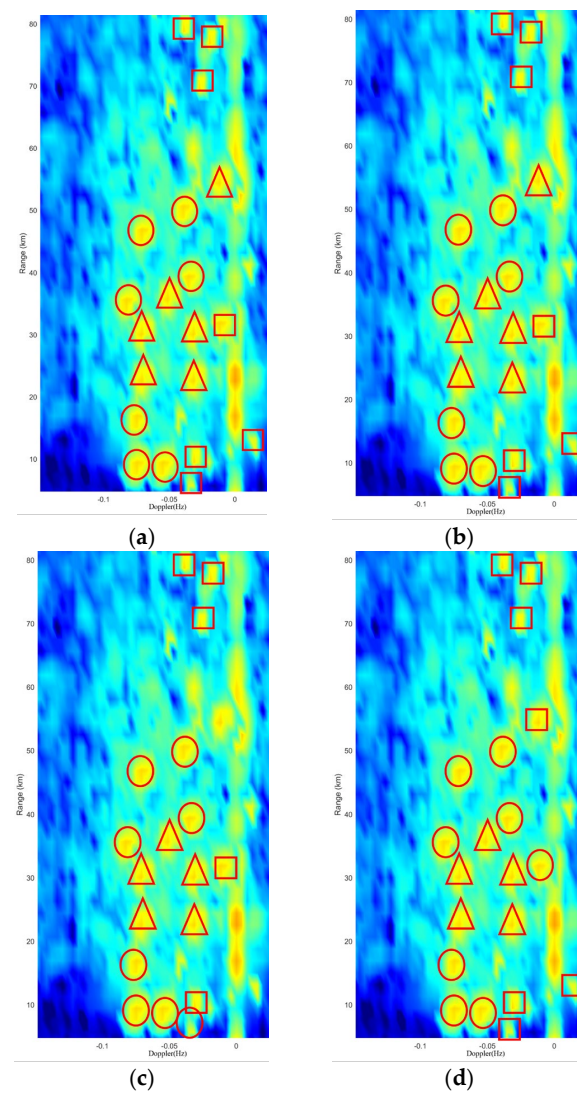

Figure 12. Targets and identification results, with the single−ship targets marked by boxes, the single−column formations marked by circles, and V−shaped formations marked by triangles: (**a**) targets in RD images; (**b**) results of the proposed cascade identification algorithm; (**c**) identification results of extremum detector−Alexnet; (**d**) identification results of extremum detector−ResNet18.

4.4. Weak Formation and Deformed Formation Identification Results

Tables 5 and 6 show the accuracy of the proposed cascade identification algorithm, extremum detector−Alexnet, and extremum detector−Resnet18. The average accuracy of the targets with SCR of 5 and 7.5 dB is around 78%. In the case of deficient target energy, deep learning networks can still extract local spatial features. The proposed algorithm achieves better accuracy than others on the same dataset. The accuracy of the three algorithms slightly decreases at the deviation of 5%. At the deviation of 20%, the shape and energy changes in the target regions cause interference with the deep learning networks, and the accuracy of the three algorithms is greatly reduced.

Table 5. Performances of the three algorithms for weak formation identification.

Dataset	Cascade Identification Algorithm	Extremum Detector−Alexnet	Extremum Detector−Resnet18
Weak set-1	88.44%	84.22%	81.67%
Weak set-2	82.11%	79.89%	74.33%

Table 6. Performances of the three algorithms for deformed formation identification.

Dataset	Cascade Identification Algorithm	Extremum Detector−Alexnet	Extremum Detector−Resnet18
Deformed set-1	93.11%	92.89%	91.33%
Deformed set-2	73.33%	73.78%	71.67%

The networks of the three algorithms are improved to meet the needs of deformable formation identification. The transfer learning can overcome the dependence on data and borrow generic features from pre-trained models for the target task [30,31]. The shallow layers of the CNN tend to extract generic features in the source domain, which could be transferred into the target domain. The features extracted by deep layers have specificity and cannot be transferred. The layer-by-layer freezing method is introduced to improve the training efficiency and robustness of the model by adjusting only the deep layer parameters. The training process is as follows:

- Label the trained CNN-ELM model as S-model;
- Freeze the first layer parameters of the S-model, adjust the parameters of the rest layers to achieve good performance for the target task, and label the trained CNN-ELM model as S-model1;
- Freeze the first two layers with fixed parameters of the S-model, retrain, and label the model as T-model2;
- Freeze the CNN parameters in the S-model, retrain the ELM, and label the model as T-model3;
- Compare T-model1, T-model2, and T-model3, and determine the model with the best performance as the transfer learning model.

The datasets "Deformed set-1" and "Deformed set-2" are split into three parts, respectively, three-twentieths of which is used as a training set, one-twentieth of which is used as a validation set, and four-fifths of which is used as a testing set.

Table 7 shows the three algorithm identification results after transfer learning. The accuracy of each model has been improved. The CNN-ELM transfer learning model on the same dataset is better than Alexnet and Resnet18 in identification accuracy.

Table 7. Performances of the three algorithms after transfer learning.

Dataset	Cascade Identification Algorithm	Extremum Detector−Alexnet	Extremum Detector−Resnet18
Deformed set-1	97.22%	96.78%	95.63%
Deformed set-2	96.22%	95.89%	95.25%

5. Conclusions

In this paper, we first investigate the spatial distribution model of ship formation for high-frequency surface wave radar (HFSWR) and propose a novel cascade identification algorithm for ship formation in the clutter edge. Taking single-column and V-shaped formations as examples, we analyze the distribution feature of the normal formation, weak formation, and deformed formation. Then, we introduce the Faster R-CNN to locate the clutter regions and propose a two-stage formation identification algorithm. In the first stage, an extremum detector based on connected regions is employed to achieve rapid detection. The proposed extremum detector reduces the number of images classified in the second stage, improving the cascade algorithm's efficiency. In the second stage, the CNN-ELM is proposed to classify the targets. Compared with the classical Alexnet and Resnet18, the proposed CNN-ELM can deal with the impact of clutter and single-ship targets well and obtain higher identification accuracy with lower computation and memory. Meanwhile, the experimental results based on the factual HFSWR background demonstrate that the proposed cascade identification algorithm is superior to the extremum detector combined with the classical CNN algorithm for ship formation identification. The proposed algorithm achieves an identification accuracy of 82% with an average signal-to-clutter ratio (SCR) of 5–7.5 dB. At the deviation of 20%, our proposed algorithm achieves an identification accuracy of 96.25% with limited samples through transfer learning. The future work will mainly focus on tracking and positioning of ship formation.

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

Data Availability Statement: The data are not accessible to the public due to part of the scientific research on the subject. Please contact the corresponding author for data.

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

References

1. Khan, R.; Gamberg, B.; Power, D.; Walsh, J.; Dawe, B.; Pearson, W.; Millan, D. Target detection and tracking with a high frequency ground wave radar. *IEEE J. Ocean. Eng.* **1994**, *19*, 540–548. [CrossRef]
2. Sun, W.; Li, X.; Pang, Z.; Ji, Y.; Dai, Y.; Huang, W. Track-to-Track Association Based on Maximum Likelihood Estimation for T/R-R Composite Compact HFSWR. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5102012. [CrossRef]
3. Green, D.; Gill, E.; Huang, W. An inversion method for extraction of wind speed from high-frequency ground-wave radar oceanic backscatter. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 3338–3346. [CrossRef]
4. Ji, Y.; Wang, Y.; Huang, W.; Sun, W.; Zhang, J.; Li, M.; Cheng, X. Vessel target echo characteristics and motion compensation for shipborne HFSWR under non-uniform linear motion. *Remote Sens.* **2021**, *13*, 2826. [CrossRef]
5. Li, H.; Shen, Y.; Liu, Y. Estimation of detection threshold in multiple ship target situations with HF ground wave radar. *J. Syst. Eng. Electron.* **2007**, *18*, 739–744.
6. Ji, Y.; Zhang, J.; Wang, Y.; Chu, X. Vessel target detection based on fusion range-Doppler image for dual-frequency high-frequency surface wave radar. *IET Radar Sonar Navig.* **2016**, *10*, 333–340. [CrossRef]
7. Ji, Y.; Liu, A.; Chen, X.; Wang, J.; Yu, C. Target Detection Method for High-Frequency Surface Wave Radar RD Spectrum Based on (VI) CFAR-CNN and Dual-Detection Maps Fusion Compensation. *Remote Sens.* **2024**, *16*, 332. [CrossRef]
8. Wang, X.; Li, Y.; Zhang, N. A robust constant false alarm rate detector based on the Bayesian estimator for the non-homogeneous Weibull clutter in HFSWR. *Digit. Signal Process.* **2020**, *106*, 102831. [CrossRef]
9. Turley, M. Hybrid CFAR techniques for HF radar. In Proceedings of the IEEE Radar Conference, London, UK, 14–16 October 1997; pp. 36–40.

10. Zebiri, K.; Mezache, A. Radar CFAR detection for multiple-targets situations for Weibull and log-normal distributed clutter. *Signal Image Video Process.* **2021**, *15*, 1671–1678. [CrossRef]

11. Hinz, J.; Holters, M.; Zolzer, U.; Gupta, A. Presegmentation-based adaptive CFAR detection for HFSWR. In Proceedings of the IEEE Radar Conference, Atlanta, GA, USA, 7 June 2012; pp. 665–670.

12. Liu, T.; Lampropoulos, G.; Fei, C. CFAR ship detection system using polarimetric data. In Proceedings of the IEEE Radar Conference, Rome, Italy, 26–30 May 2008; pp. 1–4.

13. Dzvonkovskaya, A.; Rohling, H. Adaptive thresholding for HF radar ship detection. In Proceedings of the Sixth International Radio Wave Oceanography Workshop, Hamburg, Germany, 15 May 2006.

14. Rohling, H. Radar CFAR thresholding in clutter and multiple target situations. *IEEE Trans. Aerosp. Electron. Syst.* **1983**, *4*, 608–621. [CrossRef]

15. Wang, X.; Li, Y.; Zhang, N.; Zhang, Q. CFAR Detection Based on the Nonlocal Low-Rank and Sparsity-Driven Laplacian Regularization for HFSWR. *IEEE Trans. Aerosp. Electron. Syst.* **2021**, *57*, 4472–4484. [CrossRef]

16. Li, Y.; Wu, L.; Zhang, N.; Zhang, X.; Li, Y. CFAR detection based on adaptive tight frame and weighted group-sparsity regularization for OTHR. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 2058–2079. [CrossRef]

17. Jangal, F.; Saillant, S.; Helier, M. Wavelets: A versatile tool for the high-frequency surface wave radar. In Proceedings of the IEEE Radar Conference, Waltham, MA, USA, 17–20 April 2007; pp. 497–502.

18. Baussard, A.; Grosdidier, S. Détection de cibles par radar HFSW: Utilisation des curvelets et des ondelettes continues. *Lab. EI-ENSIETA* **2009**, *1*, 116–117.

19. Grosdidier, S.; Baussard, A. Ship detection based on morphological component analysis of high-frequency surface wave radar images. *IET Radar Sonar Navig.* **2012**, *6*, 813–821. [CrossRef]

20. Li, Q.; Zhang, W.; Li, M.; Niu, J.; Wu, Q. Automatic detection of ship targets based on wavelet transform for HF surface wavelet radar. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 714–718. [CrossRef]

21. Wu, M.; Zhang, L.; Niu, J.; Wu, Q. Target detection in clutter/interference regions based on deep feature fusion for HFSWR. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 5581–5595. [CrossRef]

22. Zhang, W.; Li, Q.; Wu, Q.; Yang, Y.; Li, M. A novel ship target detection algorithm based on error self-adjustment extreme learning machine and cascade classifier. *Cogn. Comput.* **2019**, *11*, 110–124. [CrossRef]

23. Zhang, L.; You, W.; Wu, Q.; Qi, S.; Ji, Y. Deep learning-based automatic clutter/interference detection for HFSWR. *Remote Sens.* **2018**, *10*, 1517. [CrossRef]

24. Chen, Z.; He, C.; Zhao, C.; Xie, F. Enhanced target detection for HFSWR by 2-D MUSIC based on sparse recovery. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1983–1987. [CrossRef]

25. Peng, B.; Xu, J.; Xia, G.; Liu, F.; Long, T.; Yang, J.; Peng, Y. Multi-aircraft formation identification for narrowband coherent radar in a long coherent integration time. *IEEE Trans. Aerosp. Electron. Syst.* **2015**, *51*, 2121–2137. [CrossRef]

26. Liang, F.; Zhou, Y.; Li, H.; Feng, X.; Zhang, J. Multi-Aircraft Formation Recognition Method of Over-the-Horizon Radar Based on Deep Transfer Learning. *IEEE Access* **2022**, *10*, 115411–115423. [CrossRef]

27. Yang, L.; Zhang, H.; Zhou, M. Identification of ships moving in formation by HFSWR using an ISAR cross-range imaging algorithm. *Remote Sens. Lett.* **2022**, *13*, 76–86. [CrossRef]

28. Xu, G.; Xing, M.; Zhang, L.; Liu, Y.; Li, Y. Bayesian inverse synthetic aperture radar imaging. *IEEE Geosci. Remote Sens. Lett.* **2011**, *8*, 1150–1154. [CrossRef]

29. 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]

30. Zhao, C.; Guo, H.; Lu, J.; Yu, D.; Li, D.; Chen, X. ALS point cloud classification with small training data set based on transfer learning. *IEEE Geosci. Remote Sens. Lett.* **2019**, *17*, 1406–1410. [CrossRef]

31. Jing, Z.; Li, P.; Wu, B.; Yuan, S.; Chen, Y. An Adaptive Focal Loss Function Based on Transfer Learning for Few-Shot Radar Signal Intra-Pulse Modulation Classification. *Remote Sens.* **2022**, *14*, 1950. [CrossRef]