



## Article

# UltraHi-PrNet: An Ultra-High Precision Deep Learning Network for Dense Multi-Scale Target Detection in SAR Images

Zheng Zhou <sup>1</sup>, Zongyong Cui <sup>1,\*</sup>, Zhipeng Zang <sup>2</sup>, Xiangjie Meng <sup>2</sup>, Zongjie Cao <sup>1</sup> and Jianyu Yang <sup>1</sup>

<sup>1</sup> School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

<sup>2</sup> Beijing Huahang Radio Measurement Research Institute, Beijing 102445, China

\* Correspondence: zycui@uestc.edu.cn

**Abstract:** Multi-scale target detection in synthetic aperture radar (SAR) images is one of the key techniques of SAR image interpretation, which is widely used in national defense and security. However, multi-scale targets include several types. For example, targets with similar-scale, large-scale, and ultra-large-scale differences coexist in SAR images. In particular, it is difficult for existing target detection methods to detect both ultra-large-scale targets and ultra-small-scale targets in SAR images, resulting in poor detection results for these two types of targets. To solve these problems, this paper proposes an ultra-high precision deep learning network (UltraHi-PrNet) to detect dense multi-scale targets. Firstly, a novel scale transfer layer is constructed to transfer the features of targets of different scales from bottom networks to top networks, ensuring that the features of ultra-small-scale, small-scale, and medium-scale targets in SAR images can be extracted more easily. Then, a novel scale expansion layer is constructed to increase the range of the receptive field of feature extraction without increasing the feature resolution, ensuring that the features of large-scale and ultra-large-scale targets in SAR images can be extracted more easily. Next, the scale expansion layers with different expansion rates are densely connected to different stages of the backbone network, and the features of the target with ultra-large-scale differences are extracted. Finally, the classification and regression of targets were achieved based on Faster R-CNN. Based on the SAR ship detection dataset (SSDD), AIR-SARShip-1.0, high-resolution SAR ship detection dataset-2.0 (high-resolution SSDD-2.0), the SAR-ship-dataset, and the Gaofen-3 airport dataset, the experimental results showed that this method can detect similar-scale, large-scale, and ultra-large-scale targets more easily. At the same time, compared with other advanced SAR target detection methods, the proposed method can achieve higher accuracy.

**Keywords:** ultra-high precision deep learning network (UltraHi-PrNet); multi-scale; target detection; synthetic aperture radar (SAR)



**Citation:** Zhou, Z.; Cui, Z.; Zang, Z.; Meng, X.; Cao, Z.; Yang, J. UltraHi-PrNet: An Ultra-High Precision Deep Learning Network for Dense Multi-Scale Target Detection in SAR Images. *Remote Sens.* **2022**, *14*, 5596. <https://doi.org/10.3390/rs14215596>

Academic Editor: Dusan Gleich

Received: 15 September 2022

Accepted: 3 November 2022

Published: 6 November 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Synthetic aperture radar (SAR) is an active Earth observation system, which can acquire high-resolution remote sensing images. However, target detection is an indispensable part of SAR image interpretation, and its detection results directly affect the efficiency and quality of subsequent interpretation tasks [1]. In addition, target detection based on SAR images is one of the pivotal means of national battlefield reconnaissance, which can greatly enhance the capability of battlefield precision strikes, sea and land reconnaissance, military intelligence, and other tasks [2].

Fueled by SAR image technology, target detection based on SAR images has been paid more and more attention all over the world. Lincoln Studio at MIT proposed the SAR automatic target recognition (SAR-ATR) system, which focuses on target detection, target identification, and target recognition under different resolutions. They [3] also proposed that the target detection and recognition based on SAR image process follow the rule of

detection first and recognition later. Traditional SAR-ATR methods, to a great extent, rely on hand-crafted features, with much computation and poor generalization capabilities [4]. In addition, traditional target detection based on SAR image algorithms also needs to use hand-designed features for modeling, which is relatively complex [5].

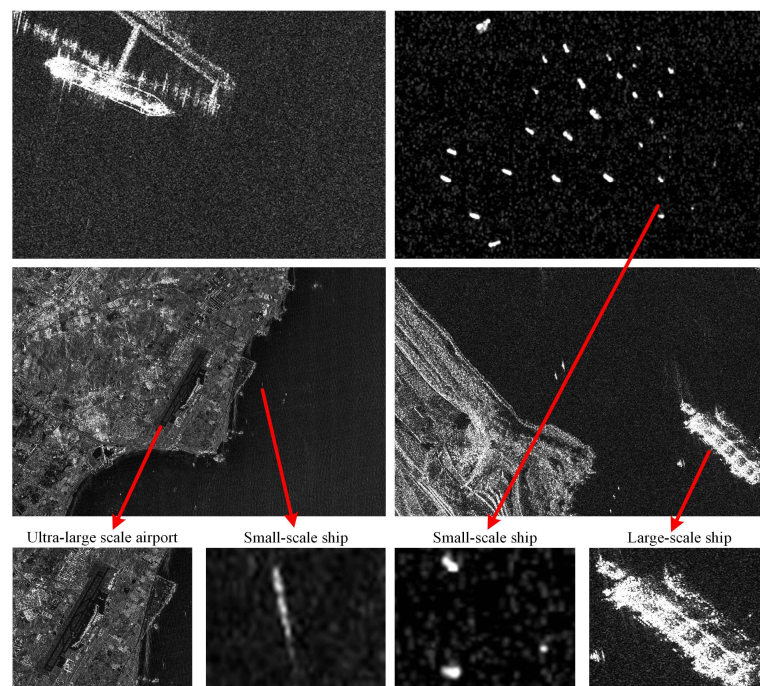
In the early stages, some scholars proposed detection methods based on segmentation, line segments, saliency, and the constant false alarm rate (CFAR) [6]. Aytekin et al. [7] and Freund et al. [8] selected the most discriminating features based on the Adaboost algorithm [9] for target detection of SAR images with complex scenes. Tang et al. [10] defined different features for line segments in SAR images and then trained support vector machines (SVMs) [11] for these features to achieve the purpose of distinguishing whether a line segment is an airport line segment and, finally, to identify each candidate region according to the classifier. Zhao et al. [12] combined the linear elementary density saliency map with enhanced learning to obtain the saliency map and, finally, obtained the saliency map for detection. Hou et al. [13] proposed multi-layer CFAR. With respect to this, firstly, they adopted a high false alarm rate, which was used to obtain the background area, and then, CFAR was conducted iteratively on the background area; finally, they detected the weak targets. He et al. [14] and Wang et al. [15] constructed a superpixel-based ship detection algorithm for polar SAR images. These methods, however, cannot realize the integration of detection and recognition and are prone to a large number of false alarms and missed detections.

In recent years, the emergence of AlexNet [16] has made convolutional neural networks (CNNs) [17] develop rapidly, and many scholars have also proposed a large number of object detection algorithms. These algorithms are rapidly becoming popular and can also detect and identify different targets simultaneously. Compared with traditional target detection methods, these algorithms do not need to design features manually, thus enhancing the generalization ability [18]. Although you only look once (YOLO) [19] can detect many targets faster, the detection effect of multi-scale targets is poor. The simultaneous single-shot multibox detector (SSD) [20] can detect small targets better, but the detection result of large targets is not ideal. To improve the detection accuracy, Region-CNN (R-CNN) [21] and its improved version Fast R-CNN [22] can detect more targets. Due to low detection efficiency, large-scale targets in SAR images cannot be detected. It can be seen that these methods are only applicable to the detection of targets of similar-scale differences.

With the increase of SAR image resolution, multi-scale target detection becomes the core problem. To solve this problem, some scholars also proposed multi-scale target detection algorithms for SAR images [23,24]. At present, multi-scale target detection algorithms in SAR images mainly include traditional algorithms and deep learning algorithms. Li et al. [25] proposed a multi-stage superpixel-based CFAR detection algorithm, which can obtain better results in simple scenes. The detection performance of this method, however, is poor when the target is in complex scenes. Zhai et al. [26] proposed a target detection algorithm with saliency and context information processing functions, which can pay more attention to large ships with prominent features and background targets, but this method ignores small ships. Hong et al. [27] proposed a YOLOv3 [28] algorithm with a multi-layer feature pyramid structure, which solved the problem of multi-scale target detection in a complex environment, but this method mainly improved the detection performance of small targets. Wang et al. [29] used attention and semantic aggregation to improve SSD [20] to detect multi-scale targets in SAR images. Although these algorithms can preliminarily solve the existing problem of multi-scale target detection, the scale of simultaneously detected targets should not be too different; otherwise, there will be missed detections.

However, in SAR images, targets with similar-scale differences, large-scale differences, and super-large-scale differences often exist at the same time, as shown in Figure 1. At the same time, we also define the targets with a large size difference and a huge size difference as cross-scale targets. In a SAR image, a target with a resolution of less than  $15 \times 15$  pixels is defined as an ultra-small target, and a target with a resolution of more than  $600 \times 600$  pixels is defined as an ultra-large target, such as airports and airplanes,

airports and vehicles, airports and ships, different types of vehicles and different types of ships, and so on. It is difficult to realize simultaneous detection of different targets, especially targets with a huge size difference. However, traditional algorithms and deep learning algorithms have poor effects on simultaneous detection. To solve this difficult problem, Lin et al. [30] made a single prediction of features extracted at different levels and achieved multi-scale target detection. Subsequently, Ren et al. [31] and Jiao et al. [32] proposed a dense-connection-based Faster R-CNN [31] algorithm that can detect more multi-scale targets. Fang et al. [33] proposed a remote sensing target detection algorithm with a pyramid structure for small targets, which improved the performance of small target detection. At the same time, Nie et al. [34] constructed a multi-scale target detection algorithm with the Mask R-CNN [35] structure, which improved the detection performance of both large and small targets. However, these methods are not sufficient to extract the features of targets with different scales and cannot simultaneously detect targets with huge size differences.



**Figure 1.** Examples of multiscale targets in SAR images, including small-scale ships, large-scale ships, and ultra-large-scale airports. There are huge differences in the scale between different targets.

Through the above analysis, both the traditional target detection algorithm and the deep-learning-based target detection algorithm have a shortcoming. When the target with large size difference appears at the same time, these methods cannot detect such targets well at the same time. Inspired by this, we designed an ultra-high precision deep learning network (UltraHi-PrNet) that can detect dense objects of different scales in SAR images. The network can achieve excellent detection performance for targets with similar-scale differences, large-scale differences, and ultra-large-scale differences in SAR images.

In conclusion, the following are the innovative parts proposed in this paper:

1. Firstly, a novel scale transfer layer is constructed, which can transfer the target features of different scales from the bottom network to the top network, while at the same time, ensuring that the ultra-small-scale and small-scale target features in SAR images can be better extracted, as well as the large-scale target features in SAR images. This method avoids the problem of missing detections of multi-scale targets in SAR images.
2. Then, a novel scale expansion layer is constructed, which can better expand the receptive field of feature extraction and can extract the features of both large-scale targets and ultra-large-scale targets simultaneously. This method solves the problem

that large-scale and ultra-large-scale targets cannot be detected simultaneously in SAR images.

3. Finally, an ultra-high precision deep learning network is established based on the ResNet101 backbone, the FPN architecture, and the Faster R-CNN [31], which can better detect ultra-small-scale targets, large-scale targets, and ultra-large-scale targets simultaneously. This method can detect targets with similar-scale differences, large-scale differences, and ultra-large-scale differences simultaneously. According to the experimental results, the algorithm has excellent performance in target detection at different scales.

The remaining sections are as follows: Section 2 describes UltraHi-PrNet and its key points in detail. Section 3 introduces the detailed contents and results of the ablation experiment. In Section 4, the results compared with other algorithms are introduced in detail. Finally, Section 5 summarizes this paper.

## 2. Proposed Method

In this section, we introduce the innovation in detail. First, we introduce the important idea of the overall structure of the proposed method. Next, the UltraHi-PrNet architecture is described in detail, including the scale transfer layer, the scale expansion layer, the RPN, and the detection network. Finally, the loss function is explained.

### 2.1. Ideas of the Method and Overall Structure

As SAR imaging technology becomes stronger and stronger, a large number of high-resolution SAR images appear, among which multi-scale target detection becomes the core problem of target detection. In some high-resolution SAR images, there are multiple objects of different scales. It is difficult for existing methods to detect similar-scale differences, large-scale differences, and ultra-large-scale differences in SAR images simultaneously.

#### 2.1.1. Ideas of the Method

In the multi-scale target detection of SAR images, it is the core of the whole process to effectively use the features of different levels to realize the detection of different scales. Therefore, in this paper, a structure with scale transfer was adopted in a bottom-up network, which can transfer the features of the lower layer to the higher layer more fully than the original network. Compared with the dense connectivity in [32], which directly adopts element addition, the proposed method can better retain the features of different scale targets. Meanwhile, the method in [32] lacks adaptive feature selection at a specific scale, which results in the disappearance of some targets at different scales.

In order to avoid the difficulty of extracting partial target features at different scales at different layers, inspired by [36], this paper injected a scale expansion layer in the process from the bottom network to the top network, which is beneficial to the network to better extract the features of the target with a larger size. In different scale layers, more target features can be extracted as much as possible. In particular, they can better extract the target features with large size differences.

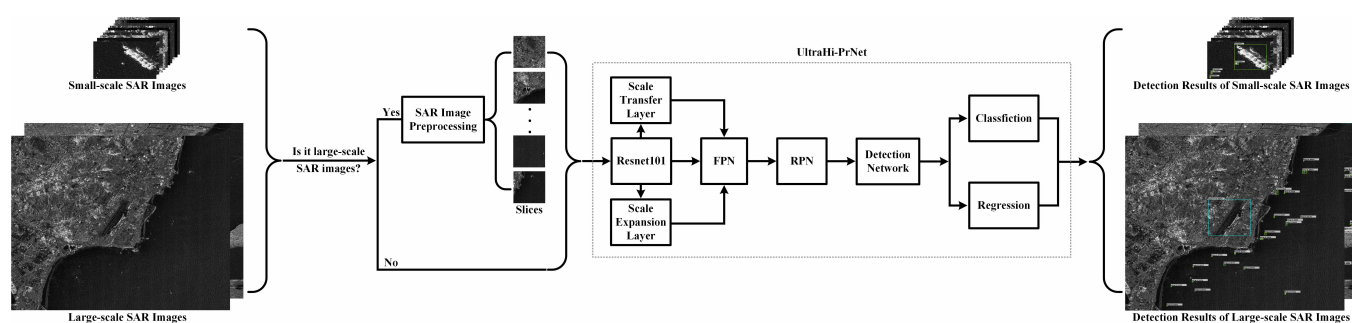
In contrast, we injected the scale transfer layer and scale expansion layer into the bottom-up network. In this way, the scale transfer layer can save more target feature maps of different scales, especially the features of small targets. While ensuring that more target features can be extracted, the scale expansion layer can more easily extract target features with larger size differences, especially features of large targets. Finally, the network can realize the effective detection of various scales of targets, with excellent precision and a low false alarm rate.

#### 2.1.2. Overall Structure

According to the above ideas, the overall process architecture of SAR image multi-scale target detection is shown in Figure 2. The whole process mainly includes the following four parts:



1. SAR image preprocessing: The size of SAR images was used to determine whether to preprocess SAR images. If the image to be detected is a large SAR image, the target occupies a small proportion of the whole large image. Therefore, the large SAR image needs to be reasonably segmented into some small images in advance, and then, target detection is carried out.
2. Feature extraction network: Firstly, the pre-processed images are fed into the backbone network for feature extraction, which is mainly composed of three parts: ResNet101, scale transfer layer, and scale expansion layer. The initial extracted features are then fed into a feature pyramid network (FPN) for feature fusion. Finally, the fused features are fed into the region proposal network (RPN) network.
3. Region proposal network: The candidate region of a multi-scale target in a SAR image is screened.
4. Detection network: The final multi-scale target detection is mainly performed by the detection head of the Faster R-CNN, including confidence scores and bounding boxes.



**Figure 2.** The entire flowchart of the algorithm. Firstly, the input SAR images of various sizes are preprocessed, then the preprocessed SAR images are transmitted to the proposed UltraHi-PrNet for feature extraction and detection, and finally, the SAR target detection results are obtained.

## 2.2. Network Architecture

Most object detection methods mainly use top-level features for prediction, such as [22,31,37], which all use the last layer of features for prediction. Those methods are only suitable for SAR images with similar-scale targets, but not suitable for SAR images with multi-scale targets. While [20,38] fused the features of different scales, the detection effect of small targets was very poor. Subsequently, the most advanced target detection method appeared [39], and the YOLOv5 methods introduced a feature pyramid network, which improved the multi-scale target detection performance; especially, it was easier to detect small-scale targets. However, small-scale targets and large-scale targets often exist simultaneously in SAR images, which cannot be detected simultaneously by simply using existing methods.

Inspired by advanced algorithms such as [30], the core network used in this paper is the feature pyramid network. Compared with other networks, it can better extract the features of multi-scale targets in SAR images and has better detection results. However, in the feature extraction stage, the network cannot well extract the features of the targets with large-scale differences in SAR images. Especially, the detection performance deteriorates when both the ultra-small-scale and ultra-large-scale targets exist in the SAR image.

To better detect targets of various scales in SAR images, we formed a scale transfer layer between each module of the pyramid network to reduce the problem of the vanishing features and vanishing gradients of small-scale targets in the process of feature extraction. In addition, To ensure that the features of large-scale targets in SAR images are retained as much as possible, we added scale expansion layers in the network process to expand the receptive field of feature extraction, so that the network is more sensitive to the features of large-scale targets and ultra-large-scale targets in SAR images.

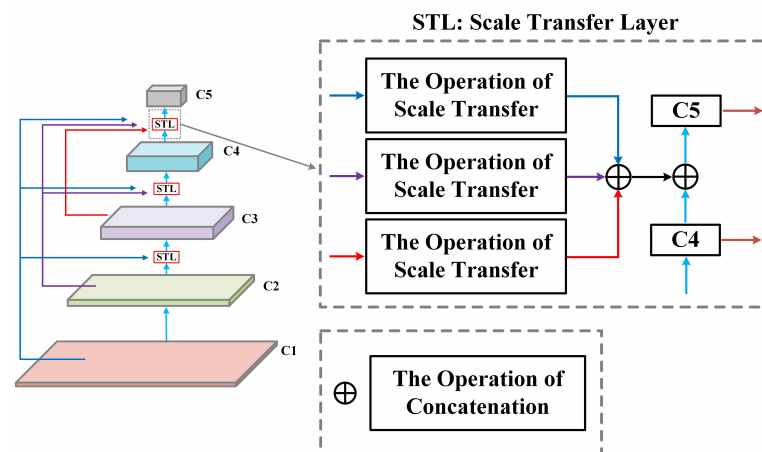
First, this article elaborates on the key components of the scale transfer layer and scale expansion layer of UltraHi-PrNet. Then, it describe each of the key parts in Figure 2.

### 2.2.1. Scale Transfer Layer

To obtain more different types of multi-scale target features in SAR images, inspired by [36,40], this paper proposes a scale transfer layer. Although [40] can make the detailed features of the target more prominent, it cannot better extract the features of targets of different scales, resulting in the failure to detect some small targets and large targets.

In the feature extraction stage, as the number of network layers increases, some target features will disappear, and the range of some feature maps will also change with the increase of network layers. The layers with no change in the feature map are regarded as a stage and use the output of this stage as the feature to be extracted in the later stage.

At this time, adding the scale transfer layer to the feature extraction stage can ensure that the low-level network does not lose the small-scale target features when extracting features, while also allowing the high-level network to more accurately extract the large-scale target features. As shown in Figure 3, we transferred all the target features of different scales in each stage to the feature map of each subsequent stage and fully retained the small-scale target features on the SAR image, so that there were small-scale target features in the feature map at each stage. As the network layers increased, there were more target features of different scales in different stages.



**Figure 3.** Structure of scale transfer layer, to realize the transition from low-level small-scale features to high-level large-scale features.

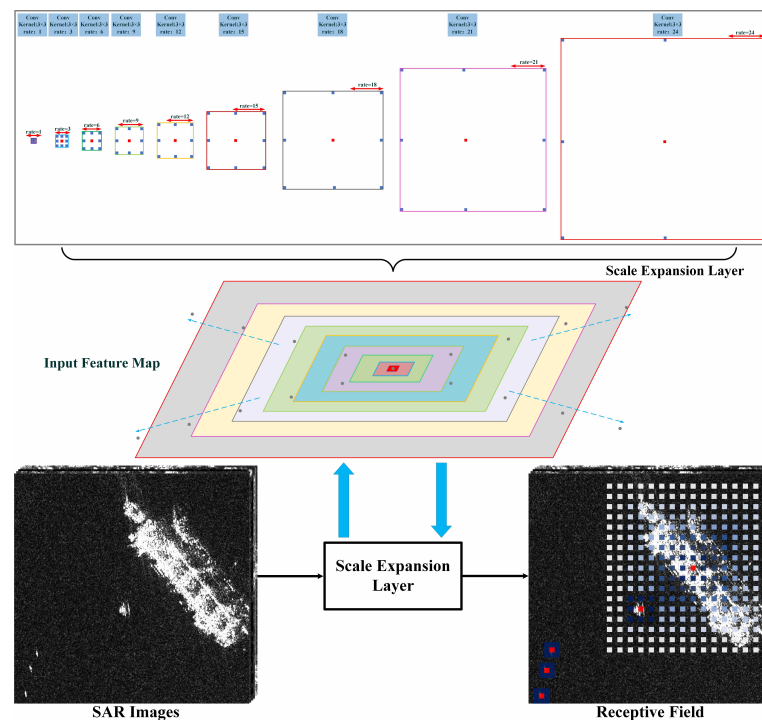
In the entire scale transfer layer, the target features of different scales output in different stages are transferred and effectively connected across stages, and the size of the feature maps in each stage is kept consistent through pooling. For example, when the target features from stage 2 Conv2 are transferred to stage 4 Conv4, the input of stage 4 Conv4 is guaranteed to be consistent in the size of feature maps of different scales in different stages. The feature scale of the target is different in different stages, and the pooling size was set as 2, while the pooling step was set as 4. Among them, for the multiple features input in each stage, we needed to effectively connect and merge them, but the feature dimension after the connection and merging would increase. In order to make the dimension the same as the input dimension of the original stage feature, we needed to add  $1 \times 1$  convolution to reduce the dimension of the connected features after the transfer to ensure that the convolutional kernel and the number of input feature channels were exactly the same.

For target detection of different scales in SAR images, the scale transfer layer can effectively improve the maximum information transfer capability between the bottom network and the top network and also alleviate the gradient disappearance. This structure retains as much as possible the target feature information of various scales in the SAR image, especially for small-scale targets and ultra-small-scale targets. This structure promotes better subsequent feature fusion effects and can also detect more different types of targets with different scales, such as small ship targets densely arranged in large SAR images.

### 2.2.2. Scale Expansion Layer

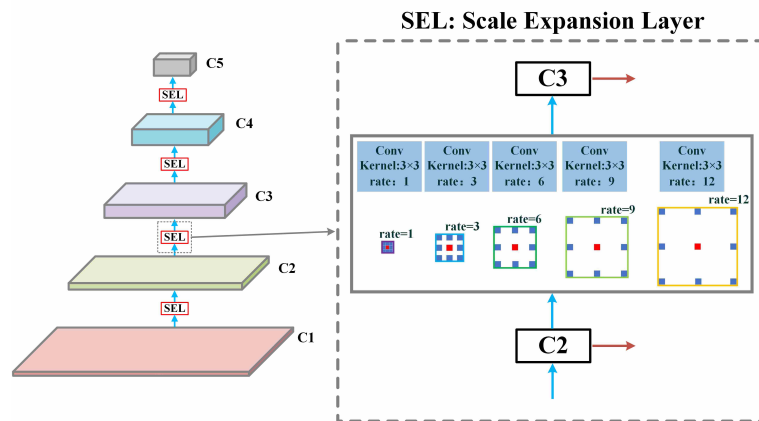
Through the scale transfer layer, we can obtain more target features of different scales, especially the features of small targets. However, in addition to small targets in SAR images, there are also targets with larger sizes. Such targets have a large difference in size, as shown in Figure 1. In the feature extraction of convolutional networks, such target features often cannot be simultaneously extracted, accompanied by feature disappearance and other problems.

In order to extract the larger-scale and ultra-larger-scale target features in the SAR image, we added a scale expansion layer in the feature extraction process, as shown in Figure 4. In addition, To ensure that the features of small and large targets in the SAR image are extracted as much as possible, we added different scale expansion layers to the network layer, as shown in Figure 5.



**Figure 4.** Structure of scale expansion layer. The feature extraction of ultra-small-scale targets and ultra-large-scale targets is realized by changing the receptive field of the scale extension layer.

The scale expansion layer includes multiple dilated convolutional kernels with different dilation rates. The data spacing is controlled by adjusting the expansion rate, and the feature extraction receptive field is increased without reducing the feature resolution, so that more useful feature representation can be learned in high-level semantic information. When the set expansion rate is larger, the corresponding convolutional kernel size is larger and the feature extraction receptive field is larger.



**Figure 5.** Architecture of dense-scale expansion layer. Different scale extension layers are placed in different network layers.

For the traditional convolutional receptive field, there is no additional expansion rate. For the  $k$ th layer, the receptive field calculation formula is as follows.

$$L_k = L_{k-1} + ((F_k - 1) * \prod_{i=1}^{k-1} S_i) \quad (1)$$

where  $L_{k-1}$  represents the receptive field size of the current convolutional layer,  $F_k$  represents the convolutional kernel size of the current convolutional layer, and  $S_i$  represents the stride.

For the scale expansion layer, the convolutional kernel and receptive field with expanded convolution are calculated as follows.

$$CKS' = (DR - 1) \times (CKS - 1) + CKS \quad (2)$$

$$RF_1 = DR \times (CKS' - 1) + 1 \quad (3)$$

$$RF_i = DR \times (CKS' - 1) + RF_{i-1} \quad (4)$$

where  $CKS$  represents the initial convolutional kernel size with a value of 3,  $DR$  represents the expansion rate,  $RF_i$  represents the receptive field of the  $i$ -th convolutional layer, and the interval between adjacent weights is  $DR - 1$ . The  $DR$  of ordinary convolution is 1 by default.

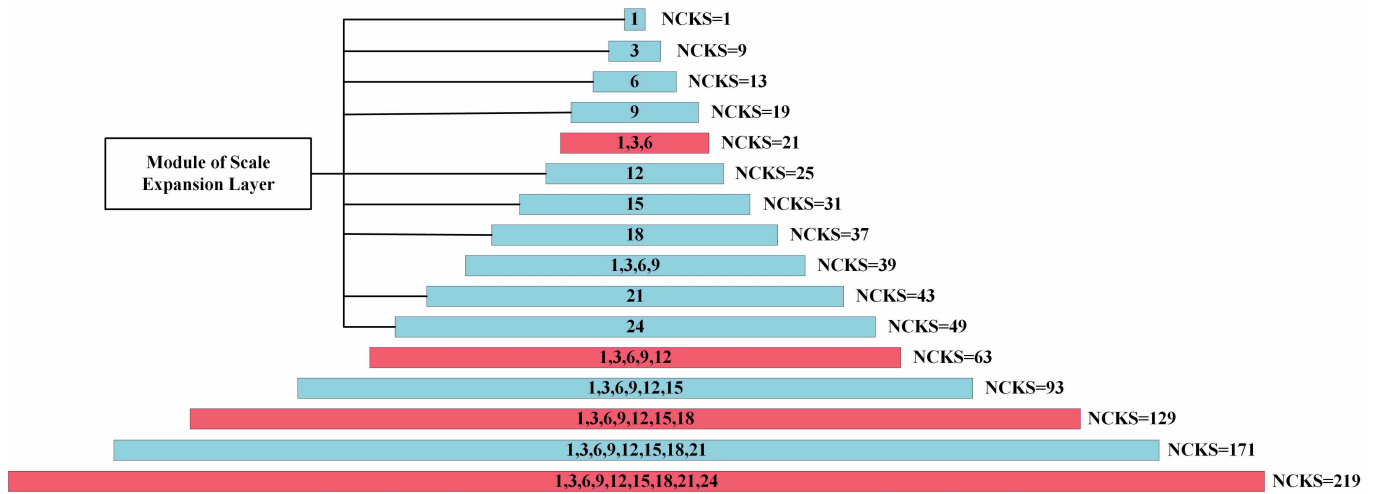
Compared with the traditional convolutional receptive field, the receptive field range of the convolutional layer with the scale expansion layer changes with the expansion rate. For example, for a  $3 \times 3$  convolutional layer, the size of the convolutional kernel is 3 when  $DR = 1$ , and the receptive field size of the first layer is 3; the size of the convolutional kernel is 55 when  $DR = 27$ , and the size of the receptive field on the first layer is 1459. It can be seen that the size of the receptive field for feature extraction will increase hundreds or even thousands of times with the increase of the expansion rate.

To obtain a larger receptive field, the scale expansion layers with different scaling rates were superimposed together to form a pyramidal dense-scale expansion layer. As shown in Figure 6, the number in each band represents the combination of different expansion rates, and the length of each band represents the convolutional kernel size of the band expansion rate after combination and is defined as  $NCKS$ . As can be seen from the figure, this structure has greater scale diversity and a greater receptive field, and it can better extract multi-scale target features from SAR images.  $NCKS$  can be represented as follows.

$$NCKS = CKS_1 + CKS_2 + \dots + CKS_i - (i - 1) \quad (5)$$



where  $CKS_i$  represents the convolutional kernel with the expansion rate in the  $i$ th layer and  $i$  is a positive integer.



**Figure 6.** Pyramidal dense-scale expansion layer. The number in the pyramid structure is the expansion rate, and the scale expansion layer marked red will be added to the network layer.

In this paper, we added the scale expansion layers of the expansion rate group  $\{1, 3, 6\}$ ,  $\{1, 3, 6, 9, 12\}$ ,  $\{1, 3, 6, 9, 12, 15, 18\}$ ,  $\{1, 3, 6, 9, 12, 15, 18, 21, 24\}$  to the layer, respectively, between  $\{C1, C2\}$ ,  $\{C2, C3\}$ ,  $\{C3, C4\}$ ,  $\{C4, C5\}$ . As shown in Figure 5, it is a gradient dense-scale pyramid network. This operation can make the network feature extraction receptive field increase tens or even thousands of times. In addition, feature images corresponding to the connection stage will be parallel to the expansion convolution in this scale expansion layer to better extract the target features of a large size in SAR images. Therefore, the proposed scale expansion layer can solve the problem that large-scale and ultra-large-scale targets in SAR images are difficult to detect.

### 2.2.3. UltraHi-PrNet

Inspired by [30,33], we designed UltraHi-PrNet as shown in Figure 7. UltraHi-PrNet can better extract low-level and high-level target features in SAR images. UltraHi-PrNet is composed of a bottom-up feedforward network, a dense-scale transfer layer, a dense-scale expansion layer, horizontal connections, and a top-down upsampling process.

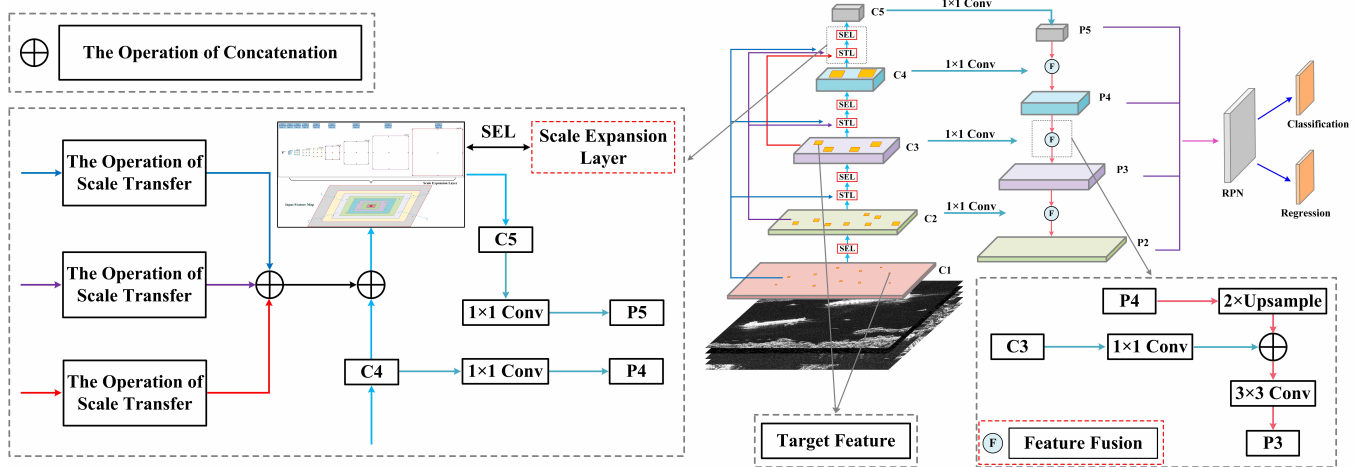
In the bottom-top process, the feedforward network mainly performs the feedforward calculation, which mainly calculates the characteristics of a multi-scale hierarchical structure. In the whole bottom-top process, the convolutional layer is composed of five stages. We chose the output from Conv1 to Conv5 as the reference set for the mapping of target features. Among the five stages, we took the final output of the last four stages as  $\{C2, C3, C4, C5\}$ , formed a scale transfer layer between the two different stages, and added a scale expansion layer between the two adjacent stages. This created UltraHi-PrNet, which is mainly used to extract target features with large-scale differences in SAR images. The calculation process is shown below.

$$CFM_n = \begin{cases} Ds(\text{Conv}_{7 \times 7}(\text{Images})), n = 1 \\ Ds(CFM_{s_{n-1}}) + \sum_{m=1}^{n-1} \text{STL}(CFM_m), n = 2, 3, 4 \\ Ds(CFM_{s_{n-1}}), n = 5 \end{cases} \quad (6)$$

$$CFM_{s_n} = CFM_n \oplus \text{SEL}(CFM_n), n = 1, 2, 3, 4 \quad (7)$$

where  $CFM_n$  is all the feature maps at the stage  $\text{Conv}_n$  in the bottom-up process,  $CFM_{s_n}$  is all the input feature maps of the stage  $\text{Conv}_{n+1}$ , STL and SEL, respectively, are the scale

transfer layer and scale expansion layer,  $D_s$  is the downsampling operation, and  $\oplus$  is the operation of concatenation.



**Figure 7.** Architecture of UltraHi-PrNet. When the scale transfer layer and scale expansion layer are placed in the network layer, the whole network can realize the extraction of various scale difference features.

In the process of top-down and horizontal linking, the top-down network upsamples the feature map with high semantic information twice and obtains the feature map with more semantic information. First,  $P_5$  is generated by  $1 \times 1$  convolution at  $C_5$ . Then, the down-top and top-down processes are fused by horizontal connection. The space size of the features to be fused is the same, and the horizontal link mainly uses a  $1 \times 1$  convolutional layer. Finally, the final feature map  $\{P_2, P_3, P_4, P_5\}$  is obtained by convolution of the fused feature map  $3 \times 3$ . Among them, a  $3 \times 3$  convolution is added to better eliminate aliasing in the upsampling process. The entire calculation formula is shown below.

$$P_n = \begin{cases} \text{Conv}_{1 \times 1}(C_n), n = 5 \\ \text{Conv}_{3 \times 3}(\text{Conv}_{1 \times 1}(C_n) \oplus f(n)), n = 2, 3, 4 \end{cases} \quad (8)$$

$$f(n) = \sum_{m=n+1}^5 \text{Upsampling}(P_m) \quad (9)$$

where  $P$  is the generated feature map, Upsampling is the upsampling operation, and  $\oplus$  is concatenation.

Therefore, we sent the final fusion feature map extracted by UltraHi-PrNet to RPN [31] to generate a bounding box proposal and used our method in Faster R-CNN [31] to achieve various scale targets for SAR image detection.

#### 2.2.4. Region Proposal Network

UltraHi-PrNet replaces the single-scale feature map without scale transfer expansion to adapt to the RPN. The whole process of the RPN is shown in Figure 8. First, the feature map obtains the intermediate layer through a  $3 \times 3$  convolutional layer. It then passes through two  $1 \times 1$  convolutional layers for classification and regression. Finally, we obtain the region proposal.  $\{P_2, P_3, P_4, P_5, P_6\}$  are respectively input into the RPN, which is used for object detection of different scales in different stages, where  $P_6$  is  $P_5$  obtained by a max pooling operation with a step size of 2.

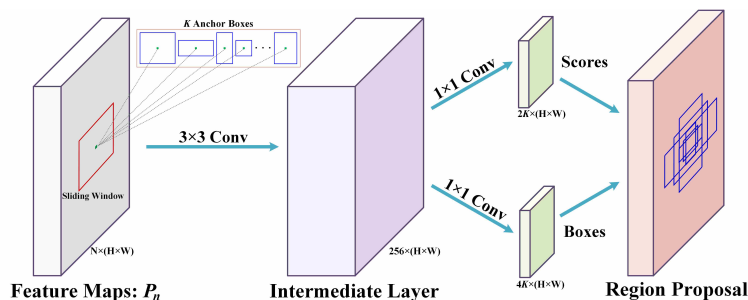


Figure 8. Architecture of the RPN.

We set  $\{16^2, 64^2, 128^2, 512^2, 1024^2\}$  different pixel regions on  $\{P_2, P_3, P_4, P_5, P_6\}$  for all anchors. We set the anchors with ratios of  $\{1 : 1, 1 : 2, 1 : 3, 2 : 1, 2 : 3, 3 : 1, 3 : 2\}$  on each stage. Therefore, there were 35 anchors on the feature map fused on UltraHi-PrNet.

### 2.2.5. Detection Network

The last part of UltraHi-PrNet is mainly composed of the region of interest (RoI) pooling and a fully connected layer, as shown in Figure 9. Firstly, in different stages of UltraHi-PrNet, the non-maximum suppression (NMS) method was used to extract the RoIs of the feature maps of different scales, and classifiers and regressors were added to the extracted RoIs. Then, RoI pooling was used to extract the target feature, and the feature was tiled onto two fully connected layers to perform classification and regression again. Finally, the detection of targets with similar-scale differences, large-scale differences, and ultra-large-scale differences in SAR images was completed.

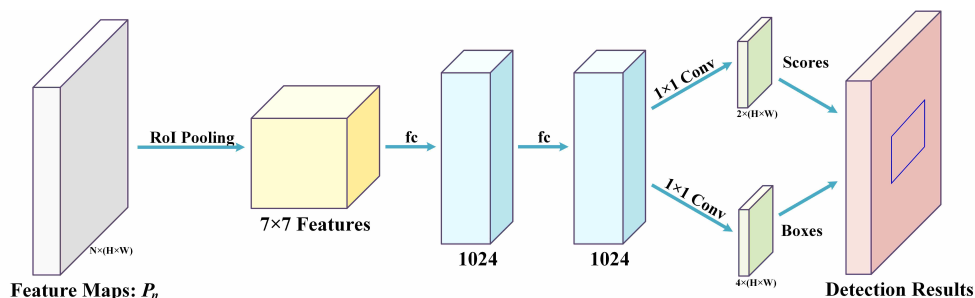


Figure 9. Architecture of the detection network.

### 2.3. Loss Function

Inspired by Faster R-CNN [31], we applied the loss function of Faster R-CNN in this paper, mainly by adding loss functions to the RPN and detection networks, which belong to multi-task loss functions. Since the outputs of both the RPN and detection networks are classification values and bounding box regression values, the loss functions in this paper included the following two types: the classification loss function was the cross-entropy loss function, and the regression loss function was the smooth L1 loss function.

## 3. Experiments and Results

In this part, first of all, we give a detailed introduction to the experimental settings, datasets, and evaluation indicators. Then, we evaluate the proposed method. We conduct performance tests in different scenarios, including simple and complex scenarios. Ablation experiments on the scale transfer layer, the scale expansion layer, and UltraHi-PrNet were conducted to verify the feasibility of the proposed algorithm. Finally, to verify the feasibility of the proposed method in large-scene SAR images, we selected large SAR images with high resolution for the experiment. The effectiveness of UltraHi-PrNet for the simultaneous detection of objects with similar-scale differences, large-scale differences, and very-large-scale differences in SAR images was verified.

### 3.1. Settings

The UltraHi-PrNet architecture proposed in this article was based on the FPN [30] architecture, and the deep learning architecture was the TensorFlow [41] system architecture. The algorithm and experiments were all verified on a computer with an Intel(R) Core(TM) i5-9400F CPU processor and an NVIDIA GeForce RTX 3060Ti GPU. The operating environment was Ubuntu, and the corresponding version was 18.04. The CUDA version was 10.2, and the corresponding cuDNN version was 7.6.5. In the experiment, ResNet-101 [42] was used as the backbone network of UltraHi-PrNet. Among them, the Resnet101 model, which was pre-trained on the ImageNet dataset [43], was selected.

### 3.2. Dataset

The datasets used in this article were: SSDD [44], AIR-SARShip-1.0, high-resolution SSDD-2.0 and SAR-ship-dataset [45], Gaofen-3 Airport Dataset. The SAR images in these datasets contain different polarization modes, resolutions, and sensors, and the airport dataset also has airport targets of different scales because of the SAR images with different resolutions. In the SAR images, airport targets near the coast are relatively larger in scale compared to ships near the coast, and there are also ultra-large and ultra-small ships close to the coast. The specific parameters are shown in Table 1.

**Table 1.** Detailed descriptions of several open SAR datasets.

Dataset	Sensor	Resolution	Polarization
SSDD	Sentinel-1, RadarSat-2	1 m–10 m	Full
AIR-SARShip-1.0	Gaofen-3	1 m, 3 m	Single
SAR-ship-dataset	Gaofen-3, Sentinel-1	5 × 5, 8 × 8, 10 × 10, etc.	Dual, Full
Gaofen-3 Airport Dataset	Gaofen-3	3 m, 5 m, 8 m, 10 m, etc.	Full

All experiments in this paper were based on the datasets proposed above. There were 1160 images selected in the SSDD dataset, 31 images selected in the AIR-SARship-1.0 dataset, 300 images selected in the high-resolution SAR ship detection dataset-2.0, 2000 images selected in the SAR-ship-dataset dataset, and 60 images selected from the dataset of Gaofen-3 airport. A total of 3551 images were counted and divided into a training set, a validation set, and a test set according to a ratio of 7:2:1. Considering that the airport dataset is a small sample and the number is very small, the 42 airport training sets were increased to 1680 through data enhancement methods such as flipping, rotating, random cropping, and brightening. From the data statistics, there were a total of 5189 datasets used for experiments, of which 4124 were training sets used for experiments, 710 verification sets, and 355 test sets.

### 3.3. Evaluation Metric

To evaluate the proposed algorithm and quantitatively evaluate the performance of the network, we used standard indicators. These mainly included: true positives (TPs), true negatives (TNs), false positives (FPs), false negatives (FNs). Recall, precision, average precision (AP), mean average precision (mAP), and other indicators were used to evaluate the superiority and feasibility of the algorithm. The calculation formula is as follows.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (10)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (11)$$

The AP is determined according to the recall and precision of each class of target. Among them, recall is the abscissa and precision the ordinate (these form the PR curve),



and AP is the area under the PR curve, where  $P$  is the precision rate of a single point and  $R$  is the recall rate of a single point.

$$AP = \int_0^1 P(R) dR \quad (12)$$

The mAP is averaged over the sum of the AP of targets in all categories.

$$mAP = \frac{\sum_{i=0}^n AP(i)}{n} \quad (13)$$

when calculating the mAP,  $AP(i)$  represents the average detection accuracy of a certain class and  $n$  represents the quantity.

### 3.4. Evaluation of UltraHi-PrNet

In this part, the performance of the algorithm is tested from three aspects: scale transfer layer, scale expansion layer, and UltraHi-PrNet, to prove the feasibility of the algorithm.

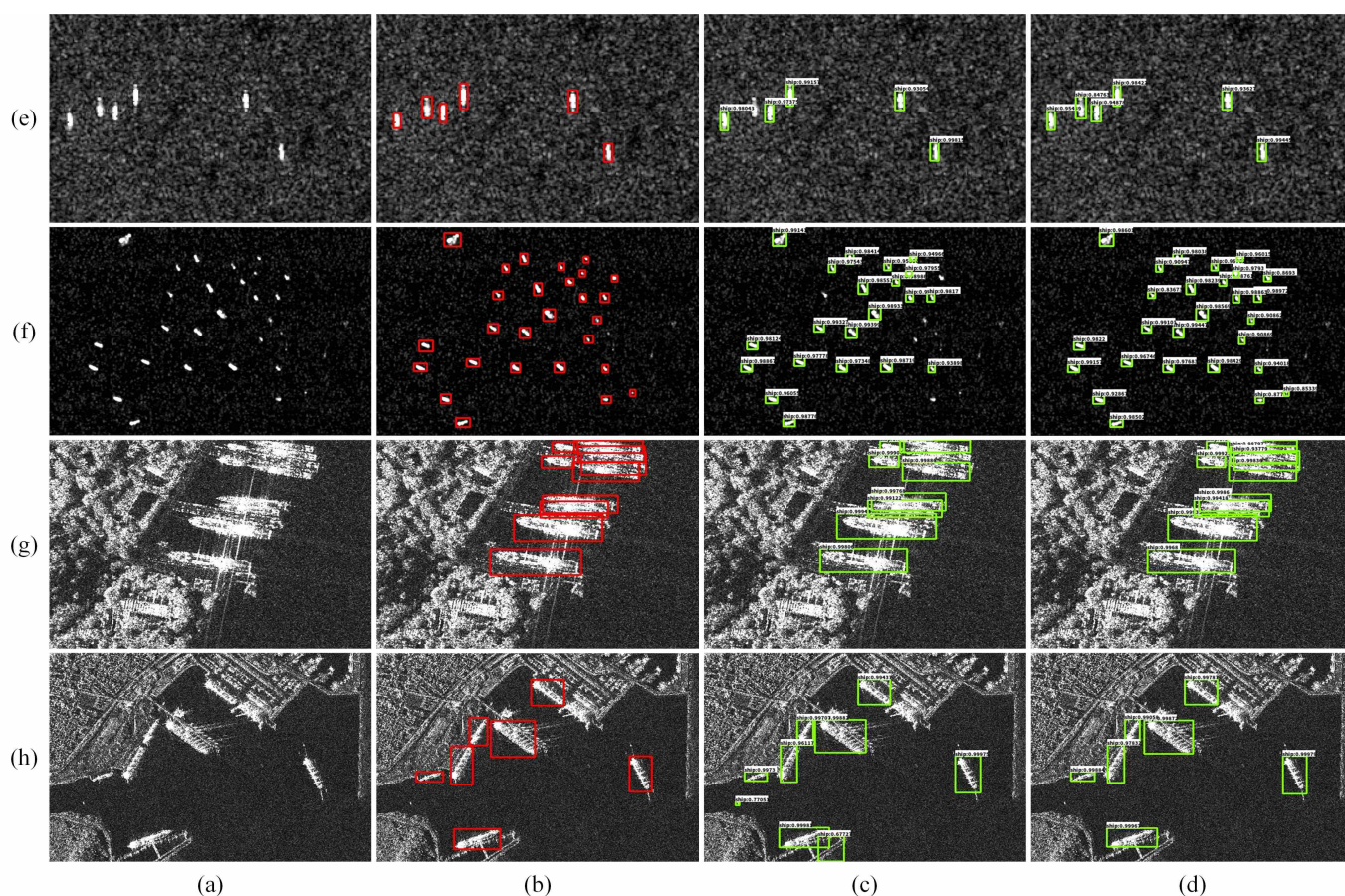
#### 3.4.1. Effect of Scale Transfer Layer

In this subsection, we perform ablation tests on the scale transfer layer. The proposed UltraHi-PrNet uses scale transfer layers to connect the feature maps of low-level networks and high-level networks. The scale transfer layer improves the information transmission ability between network layers. Moreover, the scale transfer layer obtains the feature information of various scales and extracts the features of small and medium targets and large targets in the SAR images, especially for small or extremely small targets. In addition, the structure can better extract the closely packed target feature information in SAR images.

Based on the same dataset and experimental setup, in the absence of scale expansion layers, the proposed method was verified by experiments with and without scale transfer layers. The ablation test is shown in Figure 10. The detection performance is shown in Table 2.

Figure 10a is the initial image, Figure 10b the image with the truth box, Figure 10c the detection image without the scale transfer layer method, and Figure 10d the detection image using the scale transfer layer method. Figure 10e,f are simple scenes, and Figure 10g,h are complex scenes. In simple scenes, the detection effect of the method without the scale transfer layer was obviously poor, and there were omissions of small-scale targets. In complex scenes, the detection effect of the method without the scale transfer layer was poor, and there were not only missed detections of densely arranged ship targets, but also some false alarms. The proposed method can avoid these problems. Therefore, the algorithm performed well in both simple and complex scenarios.

As shown in Table 2, the proposed algorithm had good performance in simple and complex scenarios, where the recall, precision, AP, and mAP were significantly improved, with the mAP being 4.5% higher. Because the scale transfer layer transfers the features of small targets from the bottom network to the top network, the feature information of the targets of various scales can be better preserved, and the network can more easily extract target features of different scales. As shown in Figure 10, this algorithm can achieve a better detection effect. Therefore, the scale transfer layer proposed in this paper is feasible and effective for SAR target detection. It can detect ultra-small-scale targets.



**Figure 10.** Effect of scale transfer layer. (a) is the initial image. (b) is the image with the truth box. (c) is the detection image without the scale transfer layer method. (d) is the detection image using the scale transfer layer method. As shown in the figure, (e,f) are simple scenes, and (g,h) are complex scenes.

**Table 2.** Feasibility of scale transfer layer.

Methods	Input Size	Class	Recall	Precision	AP	mAP
The original method	600 × 800	ship	91.1%	86.2%	89.5%	88.4%
		airport	90.5%	85.3%	87.3%	
The proposed method	600 × 800	ship	95.8%	89.5%	93.1%	<b>92.9%</b>
		airport	95.2%	89.6%	92.7%	

### 3.4.2. Effect of Scale Expansion Layer

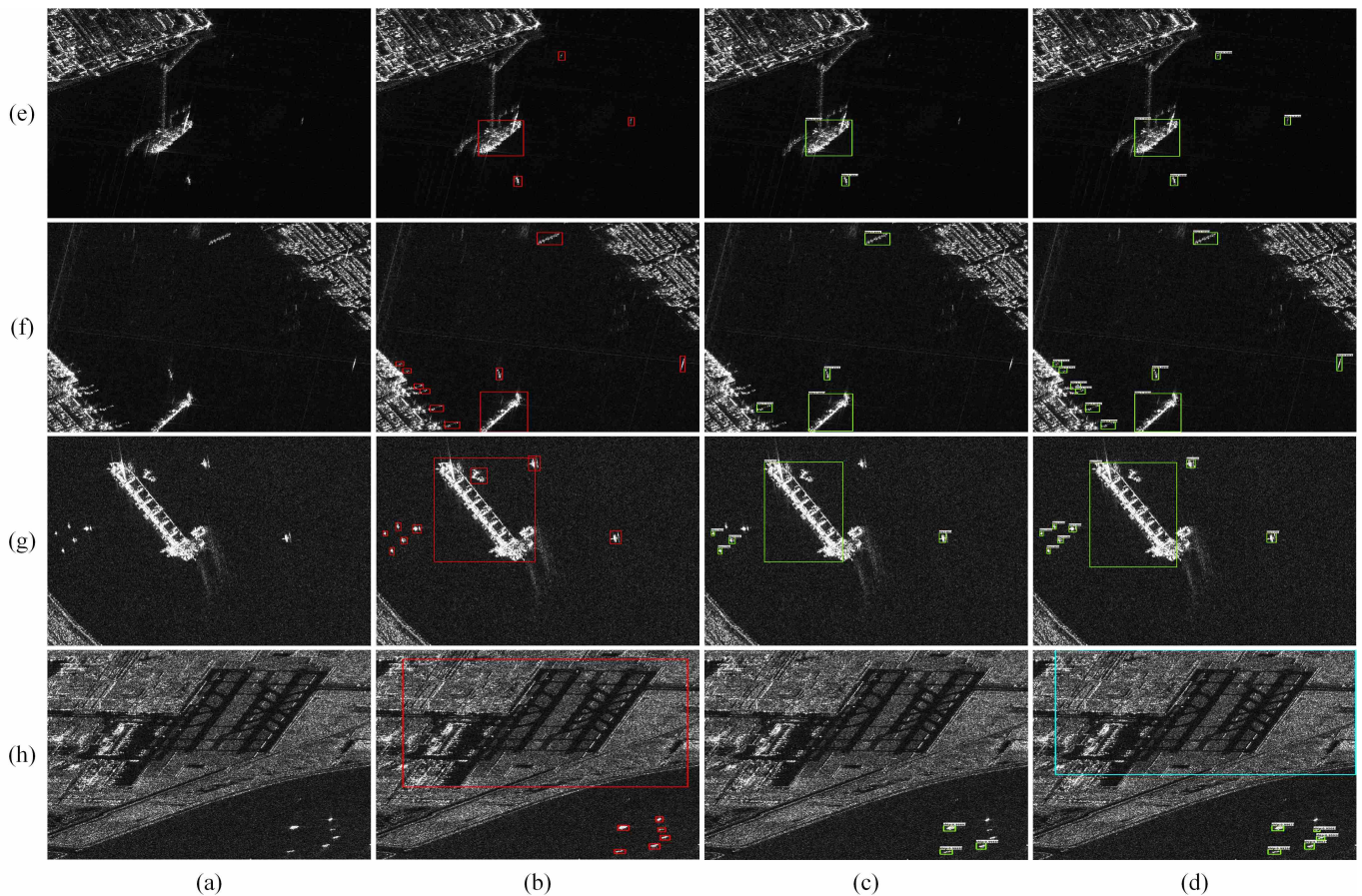
In this subsection, we perform ablation tests on the scale expansion layer. The proposed UltraHi-PrNet uses scale expansion layers to change the size of the receptive field for feature extraction. The scale expansion layers with different expansion rates were placed in different positions of the feature extraction layer. The receptive field size changes with the expansion rate. The scale expansion layer can also extract large-scale and ultra-large-scale target features from SAR images without losing the small-scale target features.

Based on the same dataset and experimental setup, in the absence of the scale transfer layer, we verified the proposed method with or without the scale expansion layer. The ablation test results are shown in Figure 11. See Table 3 for the performance comparison.

Figure 11a is the original images, Figure 11b the ground truth, Figure 11c the detection image using the scale expansion layer method, and Figure 11d the detection image without the scale expansion layer method. Figure 11e,f are simple scenes, and Figure 11g,h are complex scenes. In simple scenes, the detection effect of the scale-free expansion layer method was obviously poor, and a large number of small-scale targets were missed in the



scenes of Figure 11e,f. In complex scenes, the detection results without the scale expansion layer method were also poor, and there were also large targets missed. In Figure 11g, although large-scale ship targets were detected, many small-scale ship targets were not. In the scene of Figure 11h, the airport target cannot be detected due to its large scale. The proposed method can solve these problems in both simple and complex scenes. The algorithm can detect small-scale targets, large-scale targets, and very-large-scale targets well. Therefore, the proposed scale expansion layer is feasible and effective for SAR target detection.



**Figure 11.** Effect of scale expansion layer. (a) is the initial image. (b) is the image with the truth box. (c) is the detection image without the scale expansion layer method. (d) is the detection image using the scale expansion layer method. As shown in the figure, (e,f) are simple scenes, and (g,h) are complex scenes.

**Table 3.** Feasibility of scale expansion layer.

Methods	Input Size	Class	Recall	Precision	AP	mAP
The original method	600 × 800	ship	91.1%	86.2%	89.5%	88.4%
		airport	90.5%	85.3%	87.3%	
The proposed method	600 × 800	ship	95.4%	90.2%	92.8%	<b>92.4%</b>
		airport	95.0%	88.6%	92.0%	

As shown in Table 3. In both simple and complex scenarios, each evaluation index of the proposed method with a scale expansion layer was higher than that without a scale expansion layer. As can be seen from the table data, there was a significant increase in the recall, precision, AP, and mAP, with the mAP being 4% higher. To facilitate the network to extract more features of multi-scale targets, it is necessary to change the expansion rate so as to obtain a wider range of receptive fields. As shown in Figure 11, the proposed algorithm

can detect more targets. Therefore, the proposed scale expansion layer can greatly improve the overall detection performance and can better detect large targets and ultra-large targets.

### 3.4.3. Effect of UltraHi-PrNet

In this section, we verify the role of the UltraHi-PrNet proposed in this paper. The scale transfer layer and scale extension layer in the proposed Ultra-HI-PRNet can better extract the target features with similar-scale differences, large-scale differences, and very-large-scale differences in SAR images. It can detect not only small- and ultra-small-scale targets, but also large- and ultra-large-scale targets.

Based on the same dataset and experimental setup, we tested the proposed method with or without the scale transfer layer and scale extension layer. The ablation test is shown in Figure 12. The detection performance is shown in Table 4.

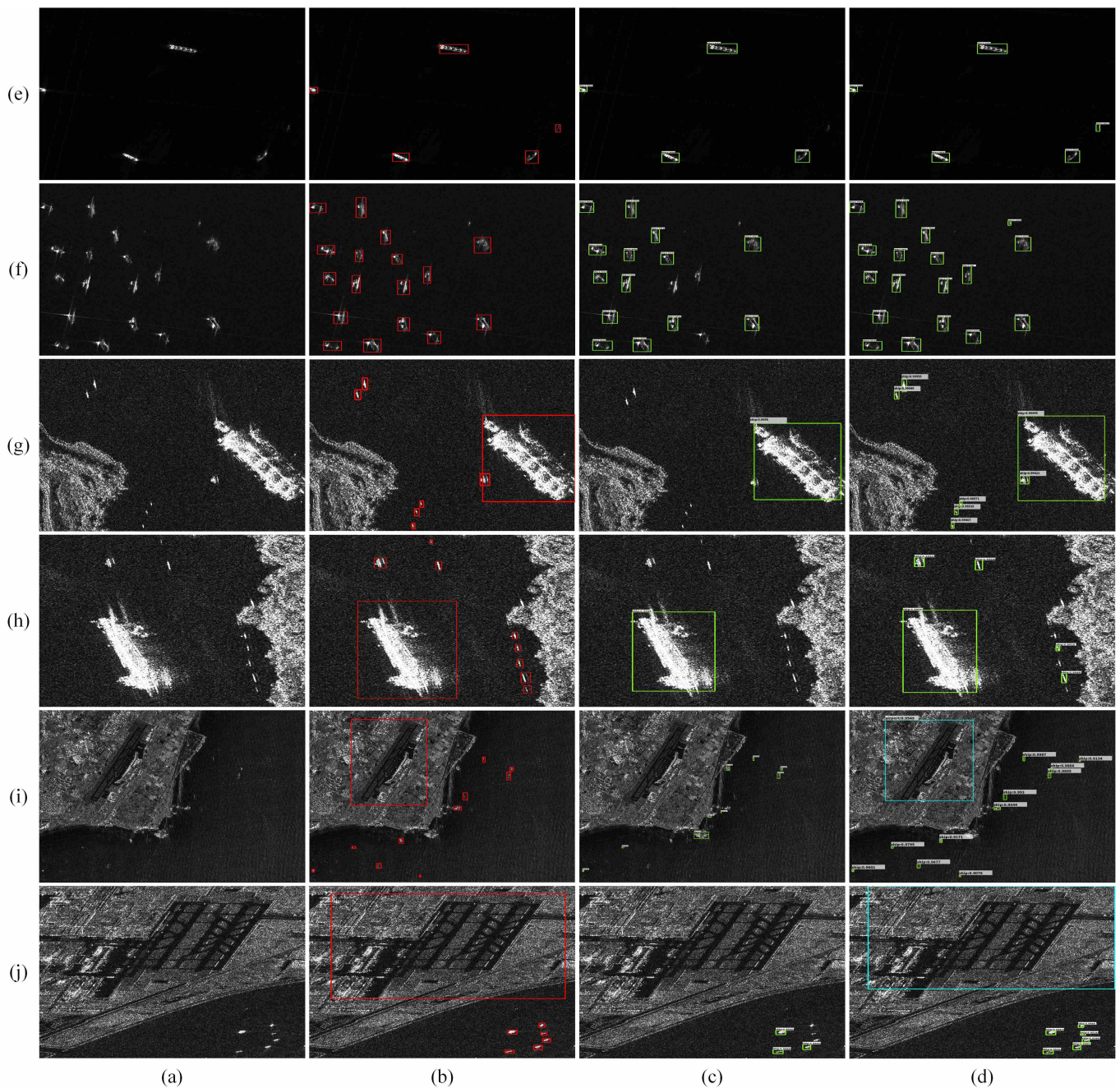
As shown in Figure 12. In simple scenes, the method without the scale transfer layer and scale expansion layer had poor results, and there were obviously small targets missed in the scene in Figure 12e,f. In complex scenes, the proposed algorithm can detect both small-scale and ultra-large-scale targets better. In the scene in Figure 12g–j, there were different targets with the ultra-large-scale difference in SAR images simultaneously. The original algorithm cannot detect the targets with large-scale differences simultaneously. Therefore, the proposed algorithm can perform well in both simple and complex scenes. It can not only realize the multi-scale target detection under general conditions, but also better detect the target with a large gap in the SAR image. Finally, the simultaneous detection of similar-scale differences, large-scale differences, and ultra-large-scale differences was realized.

As shown in Table 4, in both simple and complex scenarios, each evaluation index with a scale transfer layer and scale expansion layer was higher than that without a scale transfer layer and scale expansion layer. From the data in the table, it can be seen that the proposed method had excellent performance, in which the recall, precision, AP, and mAP all had substantial improvement, and the mAP improved by 8.5%. In this paper, the scale transfer layer can effectively guarantee that more targets in the SAR image features are extracted, and at the same time, the scale expansion layer can effectively change the scope of the receptive field of feature extraction, prompting the extracting of more multi-scale characteristics of the target, especially when small-scale targets, large-scale targets, and ultra-large-scale targets exist at the same time. Therefore, as shown in Figure 12, the UltraHi-PrNet algorithm has good performance and can effectively detect more multi-scale targets.

**Table 4.** Effect of UltraHi-PrNet.

Methods	Input Size	Class	Recall	Precision	AP	mAP
The original method	600 × 800	ship	91.1%	86.2%	89.5%	88.4%
		airport	90.5%	85.3%	87.3%	
The proposed method	600 × 800	ship	99.3%	94.8%	97.2%	<b>96.9%</b>
		airport	99.1%	93.7%	96.6%	





**Figure 12.** Effect of UltraHi-PrNet. (a) is the initial image. (b) is the image with the truth box. (c) is the result image of the method without the scale transfer layer and scale extension layer. (d) is the result image of UltraHi-PrNet. As shown in the figure, (e,f) are simple scenes, and (g–j) are complex scenes.

## 4. Discussion

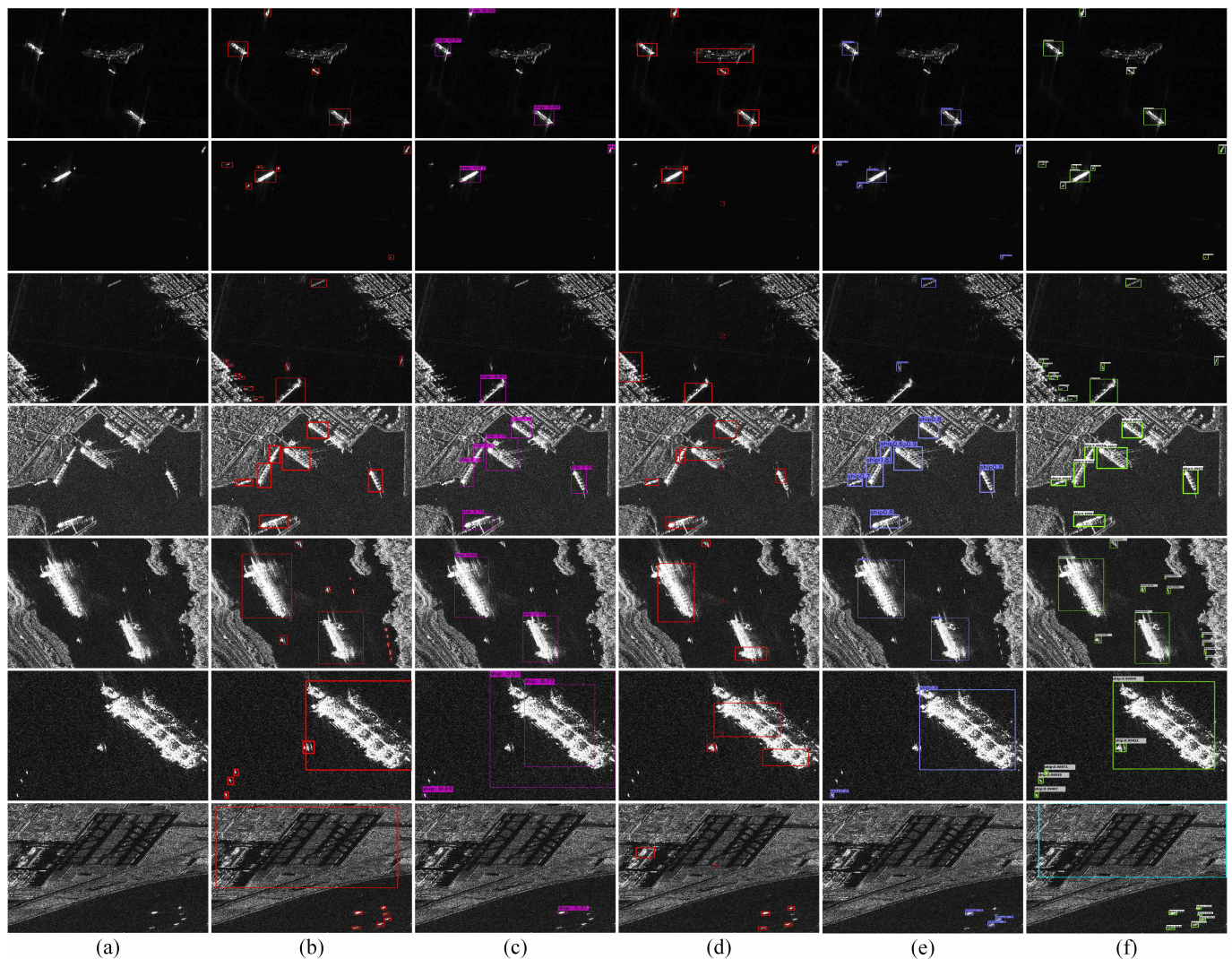
### 4.1. Comparison with Other Algorithms

For the proposed dataset, we put forward the method and the current popular method for a comparison test, and the test results were analyzed in detail. In this paper, the typical algorithms in the optical field were compared with those in the SAR field based on the FPN and CenterNet. For example, target detection methods based on one stage include SSD-300 [45], SSD-512 [45], YOLOv3 [28], YOLOv4 [39], and YOLOv5. Algorithms based on two stages include Faster R-CNN, DAPN [40], Improved Faster R-CNN [40], and PANet [46].



Center-point-based target detection methods include SSE-CenterNet [47]. All the above algorithms use the mAP for performance evaluation.

To verify whether the proposed method can simultaneously detect three types of targets with similar-scale differences, large-scale differences, and ultra-large-scale differences in SAR images, in Figure 13, we selected the above three types of SAR images in the dataset for testing. Figure 13c, Figure 13d, and Figure 13e, respectively, represent the detection images of YOLOv4, DAPN, and SSE-CenterNet. Figure 13f is the detection image of the proposed algorithm.



**Figure 13.** The proposed algorithm is compared with other advanced algorithms. (a) is the initial image. (b) is the image with the truth box. (c,d,e), respectively, represent the detection images of YOLOv4, DAPN, and SSE-CenterNet. (f) is the detection image of the proposed algorithm. As shown in the figure. There are targets with similar-scale differences, large-scale differences, and ultra-large-scale differences in the SAR images.

According to the final result of the comparative experiment (Figure 13), it can be clearly seen that the detection performance of different methods in both simple and complex scenes varied little only when the scale difference of the target was small. However, the detection performance of other methods decreased with the increase of the scale difference of the target, but the proposed algorithm could still maintain excellent detection performance. In particular, the detection performance of the other three methods was poor when there were also very-large-scale targets in the SAR image. Therefore, according to the final detection

results, the algorithm in this paper has the ability to detect small ship targets, large ship targets, and even ultra-large-scale airport targets.

On the one hand, we selected the most representative object detection algorithms in the optical field for a comparative test, for example algorithms based on YOLO4, SSD-512, Faster R-CNN, and so on. The compared detection performances are shown in Table 5.

**Table 5.** Comparison with target detection algorithms in the optical field.

Methods	Input size	Class	Recall	Precision	AP	mAP
YOLOv4	600 × 800	ship	88.9%	93.3%	88.7%	88.2%
		airport	87.9%	92.6%	87.7%	
Improved Faster R-CNN	600 × 800	ship	90.4%	87.0%	89.7%	88.8%
		airport	87.2%	83.1%	87.9%	
SSD-512	600 × 800	ship	89.8%	94.5%	89.6%	89.4%
		airport	88.1%	93.1%	89.2%	
The proposed method	600 × 800	ship	99.3%	94.8%	97.2%	<b>96.9%</b>
		airport	99.1%	93.7%	96.6%	

On the other hand, we chose the algorithms of FPN and CenterNet based on the SAR domain for a comparative test. In [40], CBAM [48] was applied to the feature pyramid network [30], so that more multi-scale targets could be detected. In [47], spatial shuffle group enhance was added to the CenterNet network [49], which improves the ability of small target detection in large SAR images. The compared detection performances are shown in Table 6.

**Table 6.** Comparison with target detection methods in the SAR field.

Methods	Input size	Class	Recall	Precision	AP	mAP
DAPN	600 × 800	ship	95.6%	90.1%	90.5%	89.8%
		airport	94.5%	88.9%	89.1%	
SSE-CenterNet	600 × 800	ship	84.2%	97.1%	95.2%	94.3%
		airport	82.6%	94.2%	93.4%	
The proposed method	600 × 800	ship	99.3%	94.8%	97.2%	<b>96.9%</b>
		airport	99.1%	93.7%	96.6%	

According to Tables 5 and 6, the proposed algorithm has good feasibility and effectiveness in comparison, where the mAP values in this paper were all higher than the mAP values of the other algorithms.

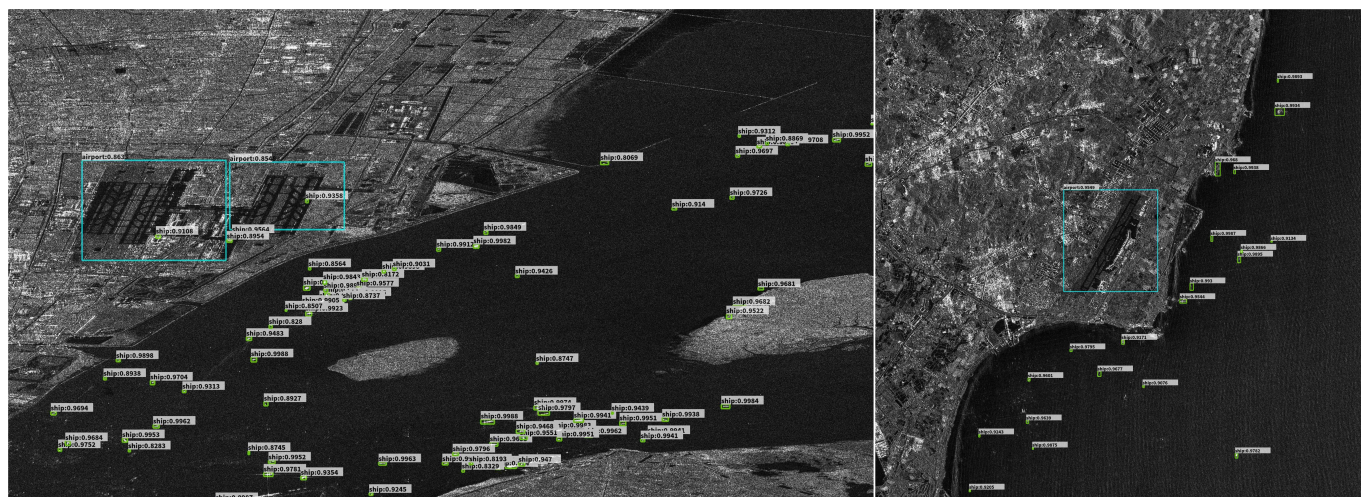
#### 4.2. Target Detection in Large-Scale SAR Images

##### Preprocessing

Due to the increasing number of SAR images, many scholars have begun to pay attention to large-scale SAR images. Since the latest methods cannot directly process the original large-sized SAR images, otherwise the detection result would be very poor and the computer would crash due to the large amount of calculation after inputting the image, we performed sliding window cropping on the large-scale SAR image according to 800 × 800 pixels with a step size of 400 pixels. The preprocessing part is shown in Figure 2.

This paper selected typical large-scene SAR images in the experimental dataset and performed detection under the same experimental settings. Figure 14 is the detection result image.





**Figure 14.** The algorithm proposed in this paper detects large SAR images. There are two kinds of targets with huge scale differences in the two ultra-large SAR images tested, which are a small ship target and an ultra-large airport target.

According to the analysis of Figure 14, the detection result was a large-scale SAR image with a complex background, and the multi-scale target in the image was composed of an ultra-large-scale airfield and a small-scale ship. Since small-scale ships account for a very small proportion in SAR images, they are not easy to detect. Meanwhile, artificial targets are also easy to identify as ships, which leads to some false alarms.

According to the result analysis, the proposed algorithm can perform well on large-scale SAR images. Because SAR targets are small and similar to speckle noise, it is easy to confuse targets and non-targets. In addition, large ships and ultra-large airports exist simultaneously with small ships in some large-scene SAR images, resulting in huge differences in target scales between different categories. At present, the most advanced algorithms still cannot detect them simultaneously, but the proposed method can still effectively detect different types of targets with large scale differences.

## 5. Conclusions

The target detection task for SAR images has important research significance and practical value. Aiming at the problem that targets with similar-scale differences, large-scale differences, and very-large-scale differences in SAR images are prone to missed detections, a deep learning network based on ultra-high precision was proposed to solve this problem.

A novel scale transfer layer was introduced into the feature extraction network, which can effectively connect the feature map of the bottom network and the top network and was beneficial to the small- and medium-scale target feature extraction from SAR images. At the same time, the scale extension layer was added after the scale transfer layer, which can change the size of the receptive field of feature extraction by adjusting the expansion rate, and the scale expansion layer with different expansion rates was connected to different stages of the feature extraction network, which can better extract the features of the ultra-large-scale target in the SAR image.

More spatial and semantic information is the prerequisite for multi-scale target detection. In order to preserve spatial and semantic information to a greater extent, the scale expansion layer and the scale transfer layer were connected effectively. In the whole feature extraction process, UltraHi-PrNet can better extract the features of small targets, as well as the features of large targets and ultra-large targets.

A large number of test results showed that the mAP value obtained by this algorithm was as high as 96.9%, and its detection performance was better than that of excellent object detection algorithms such as YOLOv4, SSE-CenterNet, DAPN, and so on. Finally, it was



verified that the algorithm can detect similar-scale differences, large-scale differences, and ultra-large-scale differences simultaneously.

**Author Contributions:** Conceptualization, Z.Z. (Zheng Zhou) and Z.C. (Zongyong Cui); methodology, Z.Z. (Zheng Zhou) and Z.C. (Zongyong Cui); software, Z.Z. (Zheng Zhou), Z.C. (Zongyong Cui) and Z.C. (Zongjie Cao); validation, Z.C. (Zongyong Cui), Z.Z. (Zhipeng Zang), X.M., Z.C. (Zongjie Cao) and J.Y.; formal analysis, Z.Z. (Zhipeng Zang) and X.M.; investigation, Z.Z. (Zheng Zhou), Z.C. (Zongyong Cui) and Z.C. (Zongjie Cao); resources, Z.C. (Zongyong Cui) and Z.C. (Zongjie Cao); data curation, Z.C. (Zongyong Cui), Z.C. (Zongjie Cao) and J.Y.; writing—original draft preparation, Z.Z. (Zheng Zhou); writing—review and editing, Z.Z. (Zheng Zhou), Z.C. (Zongyong Cui) and Z.C. (Zongjie Cao). All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China under Grant 61971101.

**Data Availability Statement:** No new data were created or analyzed in this study. Data sharing is not applicable to this article.

**Acknowledgments:** The authors would like to thank the Editors and Reviewers for the insightful and helpful comments.

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

## References

- Li, Z.; Li, S.; Liu, Z.; Yang, H.; Wu, J.; Yang, J. Bistatic Forward-Looking SAR MP-DPCA Method for Space-Time Extension Clutter Suppression. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 6565–6579. [\[CrossRef\]](#)
- Li, Z.; Zhang, X.; Yang, Q.; Xiao, Y.; An, H.; Yang, H.; Wu, J.; Yang, J. Hybrid SAR-ISAR Image Formation via Joint FrFT-WVD Processing for BFSAR Ship Target High-Resolution Imaging. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–13. [\[CrossRef\]](#)
- Novak, L.M.; Owirka, G.J.; Netishen, C.M. Performance of a high-resolution polarimetric SAR automatic target recognition system. *Linc. Lab. J.* **1993**, *6*, 11–24.
- Morgan, D.A. Deep convolutional neural networks for ATR from SAR imagery. In *Proceedings of the Algorithms for Synthetic Aperture Radar Imagery XXII*; SPIE: Bellingham, WA, USA, 2015; Volume 9475, pp. 116–128.
- Ao, W.; Xu, F.; Li, Y.; Wang, H. Detection and discrimination of ship targets in complex background from spaceborne ALOS-2 SAR images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 536–550. [\[CrossRef\]](#)
- Robey, F.C.; Fuhrmann, D.R.; Kelly, E.J.; Nitzberg, R. A CFAR adaptive matched filter detector. *IEEE Trans. Aerosp. Electron. Syst.* **1992**, *28*, 208–216. [\[CrossRef\]](#)
- Aytekin, Ö.; Zöngür, U.; Halici, U. Texture-based airport runway detection. *IEEE Geosci. Remote Sens. Lett.* **2012**, *10*, 471–475. [\[CrossRef\]](#)
- Freund, Y.; Schapire, R.; Abe, N. A short introduction to boosting. *J.-Jpn. Soc. Artif. Intell.* **1999**, *14*, 1612.
- Sun, Y.; Liu, Z.; Todorovic, S.; Li, J. Adaptive boosting for SAR automatic target recognition. *IEEE Trans. Aerosp. Electron. Syst.* **2007**, *43*, 112–125. [\[CrossRef\]](#)
- Tang, G.; Xiao, Z.; Liu, Q.; Liu, H. A novel airport detection method via line segment classification and texture classification. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 2408–2412. [\[CrossRef\]](#)
- Oliver, C.; Quegan, S. *Understanding Synthetic Aperture Radar Images*; SciTech Publishing: Raleigh, NC, USA, 2004.
- Zhao, D.; Ma, Y.; Jiang, Z.; Shi, Z. Multiresolution airport detection via hierarchical reinforcement learning saliency model. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 2855–2866. [\[CrossRef\]](#)
- Hou, B.; Chen, X.; Jiao, L. Multilayer CFAR detection of ship targets in very high resolution SAR images. *IEEE Geosci. Remote Sens. Lett.* **2014**, *12*, 811–815.
- He, J.; Wang, Y.; Liu, H.; Wang, N.; Wang, J. A novel automatic PolSAR ship detection method based on superpixel-level local information measurement. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 384–388. [\[CrossRef\]](#)
- Wang, Y.; Liu, H. PolSAR ship detection based on superpixel-level scattering mechanism distribution features. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 1780–1784. [\[CrossRef\]](#)
- Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In *Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012*; Volume 25.
- LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [\[CrossRef\]](#)
- LeCun, Y.; Bottou, L.; Bengio, Y.; Brunot, A.; Cortes, C.; Drucker, H.; Boser, B.; Henderson, D.; Guyon, I.; Sackinger, E.; et al. LeNet-5, Convolutional Neural Networks. 2015. Available online: <http://yann.lecun.com/exdb/lenet> (accessed on 14 September 2022).
- Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016*; pp. 779–788.

20. 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, Amsterdam, The Netherlands, 11–14 October 2016; pp. 21–37.
21. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Region-based convolutional networks for accurate object detection and segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *38*, 142–158. [[CrossRef](#)]
22. Girshick, R. Fast r-cnn. In Proceedings of the IEEE international Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
23. Liu, N.; Cao, Z.; Cui, Z.; Pi, Y.; Dang, S. Multi-scale proposal generation for ship detection in SAR images. *Remote Sens.* **2019**, *11*, 526. [[CrossRef](#)]
24. Dai, H.; Du, L.; Wang, Y.; Wang, Z. A modified CFAR algorithm based on object proposals for ship target detection in SAR images. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1925–1929. [[CrossRef](#)]
25. Li, T.; Liu, Z.; Xie, R.; Ran, L. An improved superpixel-level CFAR detection method for ship targets in high-resolution SAR images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *11*, 184–194. [[CrossRef](#)]
26. Zhai, L.; Li, Y.; Su, Y. Inshore ship detection via saliency and context information in high-resolution SAR images. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1870–1874. [[CrossRef](#)]
27. Hong, F.; Lu, C.H.; Liu, C.; Liu, R.R.; Wei, J. A traffic surveillance multi-scale vehicle detection object method base on encoder-decoder. *IEEE Access* **2020**, *8*, 47664–47674. [[CrossRef](#)]
28. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
29. Wang, J.; Lu, C.; Jiang, W. Simultaneous ship detection and orientation estimation in SAR images based on attention module and angle regression. *Sensors* **2018**, *18*, 2851. [[CrossRef](#)]
30. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
31. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 7–12 December 2015; Volume 28.
32. Jiao, J.; Zhang, Y.; Sun, H.; Yang, X.; Gao, X.; Hong, W.; Fu, K.; Sun, X. A densely connected end-to-end neural network for multiscale and multiscale SAR ship detection. *IEEE Access* **2018**, *6*, 20881–20892. [[CrossRef](#)]
33. Qingyun, F.; Lin, Z.; Zhaokui, W. An efficient feature pyramid network for object detection in remote sensing imagery. *IEEE Access* **2020**, *8*, 93058–93068. [[CrossRef](#)]
34. Nie, X.; Duan, M.; Ding, H.; Hu, B.; Wong, E.K. Attention mask R-CNN for ship detection and segmentation from remote sensing images. *IEEE Access* **2020**, *8*, 9325–9334. [[CrossRef](#)]
35. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
36. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 834–848. [[CrossRef](#)] [[PubMed](#)]
37. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [[CrossRef](#)]
38. Cui, Z.; Tang, C.; Cao, Z.; Dang, S. SAR Unlabeled Target Recognition Based on Updating CNN With Assistant Decision. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 1585–1589. [[CrossRef](#)]
39. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
40. Cui, Z.; Li, Q.; Cao, Z.; Liu, N. Dense attention pyramid networks for multi-scale ship detection in SAR images. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 8983–8997. [[CrossRef](#)]
41. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. {TensorFlow}: A System for {Large-Scale} Machine Learning. In Proceedings of the 12th USENIX symposium on operating systems design and implementation (OSDI 16), Savannah, GA, USA, 2–4 November 2016; pp. 265–283.
42. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.
43. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
44. Li, J.; Qu, C.; Shao, J. Ship detection in SAR images based on an improved faster R-CNN. In Proceedings of the 2017 SAR in Big Data Era: Models, Methods and Applications (BIGSAR DATA), Beijing, China, 13–14 November 2017; pp. 1–6.
45. Wang, Y.; Wang, C.; Zhang, H.; Dong, Y.; Wei, S. A SAR dataset of ship detection for deep learning under complex backgrounds. *Remote Sens.* **2019**, *11*, 765. [[CrossRef](#)]
46. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path aggregation network for instance segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 8759–8768.
47. Cui, Z.; Wang, X.; Liu, N.; Cao, Z.; Yang, J. Ship detection in large-scale SAR images via spatial shuffle-group enhance attention. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 379–391. [[CrossRef](#)]
48. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
49. Zhou, X.; Wang, D.; Krähenbühl, P. Objects as points. *arXiv* **2019**, arXiv:1904.07850.