

SAR Image Target Detection with Convolutional Neural Networks

Subjects: Engineering, Marine

Contributor: Ying Zhang, Yisheng Hao

Synthetic Aperture Radar (SAR) target detection is a significant research direction in radar information processing. Aiming at the poor robustness and low detection accuracy of traditional detection algorithms, SAR image target detection based on the Convolutional Neural Network (CNN) is presented in this entry.

Keywords: object detection ; synthetic aperture radar (SAR) ; convolutional neural network (CNN)

1. Convolutional Neural Network

1.1. Basic Theory of Convolutional Neural Network (CNN)

In today's image field, CNN is a widely used network model. The obvious distinction between this network model and the general neural network is whether it contains convolution operation. In CNN, the role of convolution operation is feature extraction, which lays the foundation for the next image processing task. There are three momentous ideas in CNN, which provide thinking for scholars to continuously improve the convolutional neural networks at different levels. These three ideas are the locally connected layer, weight sharing, and sampling layer ^[1]. These operations can improve the network's performance and lower the risk of network overfitting.

1.1.1. Locally Connected Layer

Unlike fully connected neural networks, CNN utilizes a local connection. If each pixel in the image is regarded as a neuron, then each output neuron of the fully connected network links all the neurons in the image while CNN only links a small number of adjacent neurons in space. There are two reasons for CNN to adopt local connections. Firstly, for images, local pixels are closely associated, and the correlation between pixels at farther distances is weaker. Therefore, each neuron does not need to perceive the total image but solely needs to perceive the local area. Then, the local information of low-level perception is synthesized at the high level to obtain global information. The second reason is to decrease the number of network parameters and decrease the network's complexity. It is assumed that the input image size is 200×200 , and the number of neurons in the next layer is 200. When the image is processed by full connection, the required weight parameter is $200 \times 200 \times 200 = 8 \times 10^6$. When the local connection is applied, it is presumed that the local receptive field is 5×5 , and the weight parameter is $5 \times 5 \times 200 = 5000$. So the weight parameters are reduced by 1600 times. When the input picture size is larger, the effect of lowering the number of parameters is more obvious. Local connection can also effectively avoid overfitting.

1.1.2. Weight Sharing

The values in the filter are called weights. Therefore, the so-called weight sharing refers to the convolution operation of the entire image with an identical filter. The values do not change as the position changes in the image. For CNN, the filter is generally named convolution kernel. Weight sharing is only for neurons at the same depth, and the neuron weights at different depths are not shared. Weight sharing has two functions. First, it can extract the same features at different locations in the same image. Second, it could considerably decrease the number of training parameters. For the locally connected network, the weight parameters are not shared.

1.1.3. Sampling Layer

In CNN, the layer is mainly executed by a pooling operation, so the sampling layer is also called the pooling layer. The sampling layer takes advantage of the similar local statistical characteristics of the image. The lower-level local features are aggregated into higher-level features to fully characterize the input image. The input of the sampling layer generally comes from the output of the previous convolutional layer. The sampling layer can compress the number of data and parameters, enhance the model's robustness, and reduce overfitting.

Up to the present, convolutional neural networks have evolved into numerous different structures, but their basic structures have not undergone major changes. In the basic structure, the network comprises a convolution layer, a pooling layer, and a fully connected layer.

The convolution operation is primarily to extract the image features, and feed the extracted features to the next layer for network learning. The convolution operation is completed by multiple convolution kernels. The specific procedure is to use a fixed-size convolution kernel and to traverse the entire image of the layer with a certain step size. The weight on the convolution kernel is multiplied by the corresponding position of the pixel value in the image, and then the summation operation is performed. This sum is the value after the convolution operation. After repeating the operation, a fixed-size feature map can be obtained.

The pooling operation can also be called down-sampling. The concrete process is that on the feature map after convolution, the pooling window moves in a certain order, and finally outputs an element of the feature map. Familiar pooling operations are global average pooling and maximum pooling. The pooling operation can decrease the number of parameters, speed up the program's operation, and make the model more robust as well. Common pooling window sizes are 2×2 , 3×3 , etc.

The fully connected layer is composed of some interconnected neurons, which can classify and regress the input data. The classification is to classify the target or image, and the regression is mainly to regress the parameters of the bounding box.

The nonlinear operation is to introduce some nonlinear activation functions into the network. The common nonlinear functions are sigmoid, tanh, ReLU functions, etc. [2]. In these four types of nonlinear activation functions, the ReLU function can avoid gradient vanishing and make the network better training. It can also make CNN sparse and convergence earlier. Hence, the ReLU function is the most common function in CNN.

1.2. Research Progress of CNN in Optical Image Field

The CNN can be traced back to the “neocognitron” model of Japanese scientist Fukushima, K [3]. However, due to various limitations at that time, the neural network did not attract interest. In 1998, Lecun et al. proposed the LeNet-5, which was the first time that CNN was applied to digital recognition. People began to gradually apply CNN to scientific research tasks [4]. With the introduction of nonlinear activation functions and dropout, CNN has gradually attracted people's attention. In 2012, Krizhevsky, A et al. first used CNN for large-scale image classification tasks and proposed the AlexNet, which significantly boosted the accuracy of classification. Finally, they relied on AlexNet to win first place in the ILSVRC competition [5]. AlexNet's results have sparked an upsurge in CNN research and learning. In 2014, Oxford University proposed the VGG [6], which gained runner-up in the ImageNet competition. The success of the VGG network demonstrates that increasing the network's depth could vastly increase the model's accuracy and that the VGG uses small convolution kernels (3×3 convolutional layers and 2×2 subsampling layers), which can significantly enhance the network's performance. Also in 2014, Google designed the GoogLeNet [7] and won the title in the ILSVRC2014 competition. In GoogLeNet, the most significant point is the Inception module. When constructing the network structure, the authors considers the network's depth as well as width. Under the premise that the number of parameters is decreased, the network's performance is upgraded, and the training efficiency is elevated. In 2016, He, K proposed the deep residual network called ResNet [8]. A new structure named “shortcut connections” was adopted in ResNet, which can solve the network's degradation problem, thus making it possible to train deep CNN. The number of ResNet's layers has reached 152. The accuracy rate of the image classification is 96.53%, and the recognition performance has surpassed the human eye.

In addition, some detection algorithms with superior performance have been proposed, such as two-stage target detection algorithms, which represent R-CNN [9], SPP-Net [10], Fast R-CNN [11], Faster R-CNN [12], etc. One-stage target detection algorithms include the YOLO [13] series, SSD [14], and so on. In current years, many researchers have gradually noticed that they cannot blindly upgrade the network's accuracy regardless of the number of parameters.

2. Synthetic Aperture Radar (SAR) Image Research

2.1. SAR Image Detection and Processing

2.1.1. SAR Image Dataset

For the past few years, SAR satellites have been launched all over the world, which has significantly promoted the progress of SAR image research. Based on this, many experts and scholars have constructed some SAR image data sets. These data sets contain more and more target types and image scenes, which are beneficial to the progress of SAR image research. The following five types of datasets are more common in SAR images: including the SAR ship detection dataset (SSDD), the High-Resolution-SAR Images (HRSID), the SAR-Ship-Dataset, OpenSARShip and MSTAR (Moving and Stationary Target Acquisition and Recognition).

SSDD ^[15] was constructed by Professor Li, J. The dataset contains 1160 pictures and 2456 targets, with an average of 2.12 ships per image. SSDD is the first dataset for SAR target detection. In this dataset, the images come from three different satellite sensors. The dataset has imaging pictures of four polarization modes. The resolution of the image is about 1~15m, and most of the target ships are distributed in different scenarios of nearshore or offshore. Besides, SSDD is labeled by an opensource software called Labelling, which improves considerably the accuracy of labeling ships.

HRSID ^[16] contains 5604 pictures with 16,951 targets. The size of each picture is 800 pixels × 800 pixels. The dataset uses the most advanced verifier and uses the MS COCO dataset's annotation format to make the image's label. In order to guarantee high-quality imaging, the authors chose the high-resolution imaging mode of the satellite when building the dataset. In the process of cutting the image, the offshore area and the ship-intensive area are separated, separately. For the single target in the far sea, a custom threshold is used, and 20% is used as the repetition rate of the cutting. The sliding window is 800 pixels × 800 pixels. Finally, when marking, the ship target in the image is marked in a polygonal manner, and the final file is saved in JSON format.

The SAR-Ship-Dataset ^[17] consists of two different image types: 102 GF-3 images and 108 Sentinel-1 images. The SAR-Ship-Dataset contains 43,819 ship slices, and the size of each image is 256 pixels × 256 pixels. These ships have different scales and backgrounds, which increases the target randomness. The dataset has many complex background ship targets, which provides a possibility to promote the algorithmic robustness, and can also improve the generalization algorithm performance. In the process of classifying image data, the whole dataset is randomly split into a training set, validation set and test set on the basis of the distribution ratio of 7:2:1. The image samples of the dataset are more numerous, so the target detection model can learn more abundant image features, which perfects the model's accuracy, and has a certain contribution to enhancing the model's detection performance.

OpenSARShip ^[18] is a kind of data set that comprises target types. The dataset was created by Shanghai Jiaotong University in 2017. In terms of satellite selection, the images are from the Sentinel-1A satellite. There are about a dozen types of ships in the dataset, with cargo and tanker being the largest number of them. The dataset contains approximately 10,000 SAR ship image slices. These image slices are from 41 Sentinel-1A SAR pictures. According to the different polarization methods, the slice data can be split into two categories: VH polarization and VV polarization. According to the different imaging modes, the data set can be divided into ground range detected (GRD) and single look complex (SLC). The image resolution in these two modes is 20 m × 20 m, 2.7 m × 22 m~3.5 m × 22 m, respectively. Unlike GRD mode, SLC mode also contains phase information.

The MSTAR dataset ^[19] comes from the Moving and Stationary Target Acquisition and Recognition program in the United States, which is a dataset of SAR ground stationary targets. The dataset is collected by high-resolution spotlight SAR, and the image size is 128 pixels × 128 pixels. Static SAR vehicle slices occupy the vast majority of the dataset. In the MSTAR dataset, there are not only stationary SAR vehicle images but also some environmental scene data. These scene data are obtained by SAR in strip mode, and their sizes are not exactly identical. Meanwhile, the MSTAR dataset plays a vital role in the pre-training of the model. The dominant cause is that SAR images are mostly grayscale images, which are different from RGB images. Therefore, using the MSTAR dataset for pre-training can avoid negative migration.

According to the above five types of data sets, MSTAR and OpenSARShip are data sets with target type information, and the other three are data sets without target type information. Compared with other public datasets, the MSTAR dataset has a higher resolution and an earlier time to open. Therefore, the MSTAR dataset is the most widely applied data set in SAR image target detection. The first is the sample expansion of the MSTAR dataset. Song, Q et al. ^[20] used generative adversarial networks and adversarial auto-encoders to enhance the MSTAR dataset. In addition, the related research about the MSTAR dataset also includes the improvement of CNN, the research of transfer learning, and so on. The improvement of CNN is the main research direction. In the ship target detection dataset, SSDD is a kind of data set that was published earlier. Since SSDD belongs to the data set without target category information, the research on this data set is mainly in two directions: ship target detection and target segmentation. Nowadays, most researchers use SSDD to evaluate the proposed model, which can be found in the simulation experiments in some references. In the follow-up study of SSDD, researchers also labeled the target position after rotation, making the labeled information more accurate.

However, SSDD also has a drawback, that is, the amount of data is too small. It is prone to overfitting when training the model directly by using SSDD. Therefore, in practice, it is generally used in combination with other data sets to make the model perform better. Compared with MATSR and SSDD, SAR-Ship-Dataset and HRSID were published relatively recently, so there are few studies on this data set. For the SAR-Ship-Dataset, its production team studied this dataset. Reference [21] proposed to take advantage of the RetinaNet for SAR object detection and used a feature pyramid structure to extract multi-scale features. For HRSID, Reference [22] proposed to generate simulated SAR images through sample migration and data migration, which increases the amount of the dataset and the complexity, and it was applied in SAR target detection assignments. Finally, for OpenSARship, the research focus is object recognition based on semi-supervised learning. In [23], the authors proposed a semi-supervised learning method based on a generative adversarial network, which effectively solved the over-fitting problem of complex networks caused by the small number of labeled target samples. The authors used 80%, 60%, 40%, and 20% of the labeled data in the dataset for experiments. Compared with the previous random initialization method, the results show that the accuracy is increased by 23.58%.

2.1.2. SAR Image Preprocessing

In SAR image target detection, data preprocessing is an inevitable operation, which can increase the algorithmic accuracy. The most basic operations of preprocessing are image denoising and data enhancement. Speckle noise is the main noise in SAR images, and the primary reason is the effect of the SAR imaging mechanism [24][25][26]. Therefore, it is necessary to suppress or eliminate this kind of noise for SAR image denoising. According to previous studies, the denoising algorithms can be divided into the following three aspects:

- Denoising algorithms based on spatial filtering. They mainly include Lee filtering [27], Frost filtering [28], and Non-Local-Mean (NLM) denoising [29];
- Denoising algorithms based on transform domain. They principally include wavelet domain SAR image denoising [30], shearlet domain SAR image denoising [31], and contourlet domain SAR image denoising [32];
- Recently, with the rapid progress of deep learning (DL), image-denoising algorithms based on the DL have gradually been favored by researchers. They have been diffusely applied and achieved nice results [33][34][35].

Due to the characteristics of SAR image imaging, it is often impossible to have a well-labeled large-scale SAR image dataset. For SAR image target detection, especially for CNN-based target detection algorithms, the lack of SAR image datasets is often a vital factor restricting algorithmic development. As a consequence, data expansion for small sample data sets is extremely important. Based on the existing data sets, the data sets are significantly expanded by changing the image pixels, image transformation, or noise disturbance. Finally, the neural network is trained by the expanded dataset and the original dataset simultaneously, which could greatly improve the network's performance, and increase the detection rate as well as reduce the false alarm rate.

2.2. Research on SAR Image Target Detection Based on CNN

2.2.1. Target Detection in Complex Scenes

Due to the particularity of the SAR imaging mechanism, some background clutter will inevitably occur in the imaging process. These background clutters will have an adverse influence on the SAR target detection, which will easily decrease the algorithmic accuracy and increase the false alarm rate. In response to this issue, many experts and scholars have conducted in-depth research based on the idea of CNN and achieved remarkable results. Xiao, Q proposed a multi-resolution target detection algorithm in Reference [36], which could effectively and accurately detect targets in multi-resolution SAR images, especially in complex backgrounds. For SAR image target detection in complex scenes, Yue, B [37] designed a feature extraction network based on VGG and dilated convolution, which could significantly increase the detection speed and network's accuracy. Aiming at the situation of missed detection and false detection under complex background, Xue, Y et al. [38] improved the SSD based on the knowledge of the fusion attention mechanism. Experiments show that compared with the initial SSD, the model's average accuracy is increased by 4.2%, and its anti-interference ability is also improved. In [39], in order to solve the issue of clutter interference to the detector in complex scenes, the authors proposed a SAR target detection algorithm based on a fully convolutional neural network (FCN). The core idea is to convert the target detection problem into the classification of image pixels. The test results exhibited that the algorithm could effectively decrease the false alarm target and upgrade the detection performance as well as anti-interference ability.

Offshore areas are close to land, which vastly increases the complexity of the background. Therefore, the requirements for the detection model are further improved in this case. In [40], Fu, X et al. proposed a near-shore SAR object detection

algorithm (SC-SSD) based on scene classification. The algorithm can accomplish better detection results in the case of more land scenes, and its detection speed is also significantly enhanced. Aiming at the difficulty of correctly identifying near-shore ships and land targets in the SAR images, Liu, L ^[44] proposed a new sea-land segmentation method, which used a multi-scale fully convolutional network (MS-FCN) as a foundation, and applied the target detection method based on rotating bounding box (DRBox) to offshore ship detection. Because this method combines the SAR images' global information and local information, it has high detection accuracy. Experiments show that this method can successfully locate most offshore ships. In References ^{[42][43]}, some solutions have also been proposed for target detection in complex backgrounds. Good results have been achieved.

2.2.2. Transfer Learning and Small Sample Learning Methods

It is well known that the target detection algorithm based on CNN has a relatively effective detection performance and powerful feature extraction ability. Nonetheless, the premise that the network model has this kind of ability is that it requires to be supported by a substantial amount of image data. Yet for SAR image data, it is usually challenging and costly to gain a substantial number of images, especially for SAR images with labels. Based on this, the introduction of transfer learning and small sample learning methods in SAR target detection is particularly significant.

Transfer learning refers to transferring a network to a learning task with a small amount of data after it has been fully trained on a large data set. Transfer learning is widely used in learning tasks with insufficient training data. Aiming at insufficient data in the SAR image dataset, transfer learning has been favored by many scholars. In Reference ^[44], based on the idea of fine-tuning in transfer learning, the authors first pre-trained ResNet101 on the MASAR dataset and then fine-tuned the network by using the SSDD, which effectively improved the algorithmic convergence speed and robustness of the algorithm. Based on this, Li, Y et al. ^[45] also pre-trained ResNet on the PASCAL VOC 2007 dataset and then used the pre-trained weights to initialize and fine-tune the weights. The results showed that algorithmic detection accuracy reached 94.7%, and achieved satisfactory results. Reference ^[46] adopted a two-stage transfer learning method of model fine-tuning and intra-batch balanced sampling, which effectively solved the problem of unbalanced data in SAR images.

Aiming at insufficient SAR image data, in addition to network transfer learning, data expansion is also a functional method. As mentioned above, the general dataset expansion is mainly intended for making changes on the basis of existing images, such as mirroring, translation, and flipping, for example in ^{[47][48]}. This method is relatively powerful in implementation and can play the role of data expansion. However, if a mass of data sets needs to be expanded, it is not significant to improve the model's performance by using this method alone. In the process of network training, the generator generates some images to deceive the network, and the discriminator is responsible for determining whether the data is true data, which is essentially a dynamic network game process. The use of GANs can simulate well the distribution characteristics of the original data set and can effectively enlarge the dataset. In ^[49], a GAN-based SAR image data enhancement method was proposed. This method uses a gradient penalty WGAN (Wasserstein GAN) to generate new samples based on existing SAR data, which can increase the number of samples in the training dataset. Compared with the traditional linear data generation method, the proposed method significantly improves the quantity and quality of training samples, and can effectively solve small sample recognition. In addition, Guo, Y ^[50] proposed an adaptive Faster R-CNN detection algorithm based on the knowledge of general optical images and combined it with the GAN to constrain. Simulation experiments show that this method can learn with effect and train small sample data sets, and has better performance than conventional Faster R-CNN.

2.2.3. Real-Time Model Detection and Lightweight Network

In target detecting of SAR images, some scholars are committed to improving the detection model's accuracy, but this often makes the designed or ameliorative algorithms have greater redundancy. These algorithms rely heavily on computing power, and it is difficult to achieve real-time detection requirements on the terminal. It is difficult to extensively promote it in practical applications such as real-time maritime monitoring, maritime rescue, and emergency military planning. It is not desirable to sacrifice the detection speed in order to gain higher precision. Therefore, it is necessary to further explore the detection speed and achieve a favorable trade-off between detection speed and accuracy.

To enhance the detection speed of the detection algorithm, it can be expanded from the following aspects:

One is based on the original target detection algorithm to improve. The main meliorative algorithm is based on the single-stage target detection algorithm. This is because compared with the two-stage target detection algorithm, the single-stage target detection does not extract the candidate box, which can greatly boost the detection speed. In ^[51], Chang, Y.-L proposed a real-time target detection system. The system uses YOLOv2 as a deep learning framework. In order to decrease the computational time and increase the detection accuracy, a new structure, named YOLOv2-reduced, is

developed. The detection results in SSDD and DSSDD are 90.05% and 89.13%, respectively. Compared with Faster R-CNN, the accuracy is improved and the computation time is significantly reduced. In [52], the authors also proposed an improved algorithm based on YOLOv3, which used Darknet-19 as the feature extraction backbone network and reduced the network size of traditional YOLOv3 for this specific ship detection task. The results prove that the modified algorithm has a faster detection speed and the precision is basically unchanged. References [53][54] are also based on this idea to enhance the speed of the target detection algorithm and these have attained effective results. Thus, in increasing the speed of target detection, the YOLO series algorithm is widely used.

The second is the idea of using lightweight networks. In order to make a complex network lightweight, it is typically essential to start from the following three aspects. One is to compress the trained model, such as knowledge distillation and network pruning; the second is to directly train a lightweight network, such as MobileNet and ShuffleNet; the third is improvements in hardware deployment. Based on the thinking of the knowledge distillation, a light ship detection algorithm named Tiny YOLO-Lite is proposed in [55]. The authors enhanced the channel-level sparsity of the backbone network through network pruning and used knowledge distillation to make up for performance degradation caused by network pruning. In addition, an attention mechanism was added. The simulation results reveal that although the size of the network model is only 2.8M, its detection speed has exceeded 200 fps, which significantly boosts the detection speed and upgrades the model's performance. Unlike [55], Zhou, L [56] proposed a lightweight CNN called LiraNet, which combined dense connection, residual connection, and group convolution. Based on this, a Lira--you-only-look-once (Lira-YOLO) network model was proposed, which could be easily deployed on mobile devices. The experimental results show that the complexity of the Lira-YOLO network is very low, and the number of parameters is relatively small. Simultaneously, Lira-YOLO has better detection accuracy.

In addition to the above two methods, Zhang, T [57] made full use of deep separable convolution (DS-CNN). The authors integrated a multi-scale detection mechanism, cascade mechanism, and anchor frame mechanism, and used DS-CNN instead of traditional CNN. The above operations can reduce the number of network parameters immensely and increase the detection speed. Experiments on SSDD verify the correctness and feasibility of the proposed method. Meanwhile, the network also has a strong migration generalization ability.

2.2.4. Multi-Scale Small Target Detection

In most cases, small targets in SAR images are caused by the small size or low resolution of the target itself. Since small targets occupy fewer pixels and SAR images often contain a lot of background clutter, this undoubtedly brings many difficulties to SAR image target detection. The existence of small targets in SAR images is the main factor in missed detection.

In order to enhance the performance of multi-scale small target detection, the key is to ensure that small targets will not lose information in high-level features. In CV, the common method for small target detection is feature fusion, that is to say, the location information of the underlying features and the semantic information of the high-level features are adequately fused. Only by learning the fused information can the network have better multi-scale detection ability. The idea of feature fusion is specific to the network algorithm, which is Feature Pyramid Networks (FPN) [58]. Based on YOLOv3, Hu, C et al. [48] introduced the design idea of a residual network as well as a feature pyramid structure and also introduced a class of balance factors, which can effectively optimize the weight of small targets in the loss function. The results demonstrate that the algorithm has better detection performance for small targets. Reference [59] introduced the attention mechanism into the target detection algorithm of multi-scale feature fusion, which significantly improved the detection performance of small targets. In order to solve multi-scale target detection in spaceborne SAR images, Liu, S et al. [60] put forward a new detector called Receptive Field Block (RPF). RPF adds dilated convolution and uses four residual structures to connect the input and output of the branch. In addition, the authors also thoroughly considered the effect of the parameters on the model's performance, replacing the original 7×7 convolutions with 1×7 and 7×1 convolutions, which significantly decreases the model's complexity. Experiments on the SSDD reveal that the model's mAP reaches 95.03%. The detection speed increased to 47.16 FPS, and the model size also decreased significantly. Aiming at the poor sensitivity of the model to different ship scales in ship detection, Cui, Z et al. [61] proposed a dense attention pyramid network (DAPN) based on the FPN. The structure makes full use of the CBAM module to completely connect the bottom and top features of the feature pyramid. This method extracts rich features containing resolution and semantic information and solves the problem of multi-scale ship detection. The simulation results show that this method has extremely high detection precision, but the model has poor adaptability to different scenarios. Further improvement and research are needed for this problem.

2.2.5. Combination of Traditional Target Detection Algorithm and CNN

In the traditional SAR target detection algorithm, CFAR is the earliest and most mature. Its computational complexity is also lower than that of CNN. Therefore, the combination of CFAR and CNN can decrease the number of parameters and increase the detection speed. Reference [62] proposed a ship-borne detection method based on traditional CFAR and lightweight DL. The experimental results show that the detection algorithm has a high detection speed and basically achieves the effect of real-time detection. Cui, Z et al. [63] proposed a constant false alarm rate (CP-CFAR) detection algorithm with convolution and pooling. The convolution layer in this algorithm uses horizontal and vertical Sobel operators to improve the contrast between the target and the background. The pooling layer reduces the processing dimension of the image. Adding a convolution layer and a pooling layer before a two-parameter CFAR can lessen computational elements without losing the main features of the original image. The simulation results show that the algorithm has fast detection speed and the running time is less than 192ms. Aiming at low detection accuracy of multi-scale ship targets, a target detection algorithm combining Faster R-CNN and CFAR was proposed in [64]. The algorithm uses Faster R-CNN to generate different sizes of regional proposal boxes. For some low-confidence proposal boxes, let CFAR detect them. Finally, the detection results with high confidence and the CFAR detection results are taken as the final output. The simulation results demonstrate that the algorithm can effectively resolve the problem of multi-scale target detection. However, for some small targets, its detection performance is poor, so the algorithm needs to be further improved for this problem.

Based on the above literature analysis, it can be concluded that the dominating idea of CNN-based SAR object detection is to improve the original CV algorithm. In allusion to different problems, different algorithms need to be applied. For SAR target detection in complex scenes, the difficulty is the interference of the scene clutter to target detection. The existence of background clutter makes the false alarm rate higher, which requires that the algorithm should have better robustness. For the detection of dense small targets, feature fusion is one of the more feasible schemes. When detecting small targets in SAR images, the down-sampling of CNNs will cause the loss of some small target information, which cannot be transmitted to the deep neural network. Due to the existence of feature fusion, the underlying small target information will be better fused with the deep semantic information. As a result, the network can learn more feature information about the image. Of course, for small target detection, not only features fuse but also other methods can be used. The deepening of network layers is conducive to obtaining detection information of small targets. Therefore, the width of CNN can be increased while the network depth can be reduced, such as the Inception module. This is also a more feasible solution. To increase the model's detection speed, the key is to decrease the parameters of the model, but the prerequisite is that the model's detection accuracy cannot be reduced too much. Real-time detection of SAR image targets is also very important, which requires reducing the complexity of the model so that the model can be effectively operated at the device terminal. Unlike ordinary RGB optical image target detection, the lack of effective data sets in SAR images is one of the significant factors that hinders the application of CNNs in this field.

2.3. Summary of Research Status

Based on the above literature analysis, the four target detection algorithms have their own merits. The target detection algorithm based on structural features, compared with the other three types of target detection algorithms, has better stability, stronger robustness, and a slight advantage in detection speed. However, its apparent deficiency is the need for prior information. It can be said that the correctness of prior information determines the subsequent detection effect. Therefore, such algorithms rely heavily on prior information. Meanwhile, in some complex backgrounds, especially in near-shore target detection, this algorithmic detection accuracy is often relatively low. The target detection algorithm based on gray features has the outstanding advantage of effective stability and is not easily interfered by background clutter. Nevertheless, similar to the detection algorithm based on structural features, this kind of algorithm still needs prior information, which will limit its large-scale promotion and use of algorithms. At the same time, this kind of algorithm has difficulty in establishing a unified target statistical model and hard to achieve real-time processing in the face of a large number of image data. The detection efficiency is also low. The target detection algorithm based on texture features, compared with the first two detection algorithms, has higher detection accuracy, but this kind of algorithm is time-consuming when dealing with texture feature extraction, so the timeliness is poor. Moreover, when some dense targets are detected (such as the detection of dense ships at sea), the targets are dense and the distance between the targets is relatively close, so this will affect the calculation of the extended fractal, and then some targets will be missed. Therefore, this kind of algorithm is not suitable for detecting dense small targets. Finally, the target detection algorithm based on CNN is analyzed. The target detection algorithm based on CNN has the advantages of high accuracy and fast detection speed. However, its shortcomings are also obvious. The premise of CNN with such excellent characteristics is that it needs a substantial number of image data samples to train the network. Insufficient data is rarely seen in ordinary RGB optical image target detection. However, in the field of SAR images, based on the previous analysis, it is challenging to obtain a labeled SAR image data set with a mass of data. Therefore, in SAR target detection, insufficient data is a key factor that restricts the large-scale use of CNN in the SAR field. Meanwhile, the detection algorithm based on CNN has

higher requirements for hardware GPU devices. Without a decent device, it is difficult to train a better CNN model. Therefore, the target detection algorithm using CNN should consider this limitation. The full and reasonable application of CNN in SAR image target detection still requires further research.

References

1. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016; pp. 331–360.
2. Wang, H.; Wang, Y.; Lou, Y.; Song, Z. The role of activation function in CNN. In *Proceedings of the 2020 2nd International Conference on Information Technology and Computer Application (ITCA)*, Guangzhou, China, 1 December 2020; p. 429–432.
3. Fukushima, K. Neocognitron: A Self-Organizing Neural Network Model for A Mechanism of Pattern Recognition Unaffected by Shift in Position. *Biol. Cybern.* 1980, 36, 193–202.
4. Lecun, Y. Gradient-Based Learning Applied to Document Recognition. *Proc. IEEE* 1998, 86, 2278–2324.
5. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* 2017, 60, 84–90.
6. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. In *Proceedings of the International Conference on Learning Representations*, San Diego, CA, USA, 7–9 May 2015; pp. 1–14.
7. Zeiler, M.D.; Fergus, R. Visualizing and understanding convolutional networks. In *Proceedings of the 2014 European Conference on Computer Vision*, Zurich, Switzerland, 5–12 September 2014; pp. 818–833.
8. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.
9. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
10. 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, 1901–1916.
11. Girshick, R. Fast R-CNN. *arXiv* 2015, arXiv:Abs/1504.08083.
12. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans.* 2017, 39, 1137–1149.
13. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 779–788.
14. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Berg, A.C. SSD: Single shot multibox detector. In *Proceedings of the 2016 European Conference on Computer Vision*, Amsterdam, Netherlands, 11–14 October 2016; pp. 21–37.
15. 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*, Beijing, China, 13–14 November 2017; pp. 1–6.
16. Wei, S.; Zeng, X.; Qu, Q.; Wang, M.; Su, H.; Shi, J. HRSID: A High-Resolution SAR Images Dataset for Ship Detection and Instance Segmentation. *IEEE Access* 2020, 8, 120234–120254.
17. Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common Objects in Context. In *Proceedings of the 2014 European Conference on Computer Vision*, Zurich, Switzerland, 5–12 September 2014; pp. 740–755.
18. Huang, L.; Liu, B.; Li, B.; Guo, W.; Yu, W.; Zhang, Z.; Yu, W. OpenSARShip: A Dataset Dedicated to Sentinel-1 Ship Interpretation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 2018, 11, 195–208.
19. Hummel, R. Model-based ATR using synthetic aperture radar. In *Proceedings of the Record of the IEEE 2000 International Radar Conference*, Alexandria, VA, USA, 1 February 2000; pp. 856–861.
20. Song, Q.; Xu, F.; Zhu, X.X.; Jin, Y.Q. Learning to Generate SAR Images with Adversarial Autoencoder. *IEEE Trans. Geosci. Remote Sens.* 2022, 60, 1–15.
21. Wang, Y.; Wang, C.; Zhang, H.; Dong, Y.; Wei, S. Automatic Ship Detection Based on RetinaNet Using Multi-Resolution Gaofen-3 Imagery. *Remote Sens.* 2019, 11, 531.

22. Zhu, C.; Zhao, D.; Qi, J.; Qi, X.; Shi, Z. Cross-domain transfer for ship instance segmentation in SAR images. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium, Brussels, Belgium, 11 July 2021; pp. 2206–2209.
23. Zhang, W.; Zhu, Y.; Fu, Q. Semi-Supervised Deep Transfer Learning-Based on Adversarial Feature Learning for Label Limited SAR Target Recognition. *IEEE Access* 2019, 7, 152412–152420.
24. Lu, Z.; Jia, X.; Zhu, W.; Zeng, C. Study on SAR Image Despeckling Algorithm. *J. Ordnance Equip. Eng.* 2017, 38, 104–108.
25. Goodman, J.W. Some Fundamental Properties of Speckle. *J. Opt. Soc. Am.* 1976, 66, 1145–1150.
26. Eom, K.B. Anisotropic Adaptive Filtering for Speckle Reduction in Synthetic Aperture Radar Images. *Opt. Eng.* 2011, 50, 97–108.
27. Lee, J.S.; Jurkevich, L.; Dewaele, P.; Wambacq, P.; Oosterlinck, A. Speckle Filtering of Synthetic Aperture Radar Images: A Review. *Remote Sens. Rev.* 1994, 8, 313–340.
28. Frost, V.S.; Stiles, J.A.; Shanmugan, K.S. A Model for Radar Image and Its Application to Adaptive Digital Filtering of Multiplicative Noise. *IEEE Trans. Pattern Anal. Mach. Intell.* 1982, 4, 157–165.
29. Torres, L.; Frery, A.C. SAR Image Despeckling Algorithms Using Stochastic Distances and Nonlocal Means. *arXiv* 2013, arXiv:Abs/1308.4338.
30. Liu, S.; Hu, S.; Xiao, Y. SAR Image Denoising Based on Wavelet Contourlet Transform and Cycle Spinning. *Signal Process.* 2011, 27, 837–842.
31. Liu, S.; Shi, M.; Hu, S.; Xiao, Y. Synthetic Aperture Radar Image De-Noising Based on Shearlet Transform Using the Context-Based Model. *Phys. Commun.* 2014, 13, 221–229.
32. Fang, J.; Wang, D.; Xiao, Y.; Ajay Saikrishna, D. Denoising of SAR images based on wavelet-contourlet domain and PCA. In Proceedings of the 2014 12th International Conference on Signal Processing (ICSP), Hangzhou, China, 19–23 October 2014; pp. 942–945.
33. Zhang, K.; Zuo, W.; Gu, S.; Zhang, L. Learning deep CNN denoiser prior for image restoration. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2808–2817.
34. Lv, B.; Zhao, J. Image Denoising Algorithm Based on Composite Convolutional Neural Network. *Pattern Recognit. Artif. Intell.* 2017, 30, 97–105.
35. Wang, P.; Zhang, H.; Patel, V.M. SAR Image Despeckling Using a Convolutional Neural Network. *IEEE Signal Process. Lett.* 2017, 24, 1763–1767.
36. Xiao, Q.; Cheng, Y.; Xiao, M.; Zhang, J.; Shi, H.; Niu, L.; Ge, C.; Lang, H. Improved Region Convolutional Neural Network for Ship Detection in Multiresolution Synthetic Aperture Radar Images. *Concurr. Comput. Pract. Exper.* 2020, 32, 1–10.
37. Yue, B.; Han, S. A SAR Ship Detection Method Based on Improved Faster R-CNN. *Comput. Mod.* 2019, 9, 90–101.
38. Xue, Y.; Jin, G.; Hou, X.; Tan, L.; Xu, J. SAR Ship Detection Method Incorporating Attention Mechanism and Improved SSD Algorithm. *Comput. Appl. Res.* 2022, 39, 265–269.
39. Zhang, Y.; Zhu, W.; Wu, X. Target Detection Based on Fully Convolutional Neural Network for SAR Images. *Telecommun. Eng.* 2018, 58, 1244–1251.
40. Fu, X.; Wang, Z. SAR Ship Target Rapid Detection Method Combined with Scene Classification in the Inshore Region. *J. Signal Process.* 2020, 36, 2123–2130.
41. Liu, L.; Chen, G.; Pan, Z.; Lei, B.; An, Q. Inshore ship detection in SAR images based on deep neural networks. In Proceedings of the 2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 1 July 2018; pp. 25–28.
42. Dai, W.; Mao, Y.; Yuan, R.; Liu, Y.; Pu, X.; Li, C. A Novel Detector Based on Convolution Neural Networks for Multiscale SAR Ship Detection in Complex Background. *Sensors* 2020, 20, 2547.
43. Fu, J.; Sun, X.; Wang, Z.; Fu, K. An Anchor-Free Method Based on Feature Balancing and Refinement Network for Multiscale Ship Detection in SAR images. *IEEE Trans. Geosci. Remote Sens.* 2021, 59, 1331–1344.
44. Liu, J.; Zhao, T.; Liu, M. SAR Ship Target Rapid Detection Method Combined with Scene Classification in the Inshore Region. *J. Hunan Univ.* 2020, 47, 85–91.
45. Li, Y.; Ding, Z.; Zhang, C.; Wang, Y.; Chen, J. SAR ship detection based on resnet and transfer learning. In Proceedings of the 2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019.

46. Shao, J.; Qu, C. CNN Based Ship Target Recognition of Imbalanced SAR Image. *Electron. Opt. Control* 2019, 26, 90–97.
47. Wang, Z.; Du, L.; Mao, J.; Liu, B.; Yang, D. SAR Target Detection Based on SSD with Data Augmentation and Transfer Learning. *IEEE Geosci. Remote Sens. Lett.* 2019, 16, 150–154.
48. Hu, C.; Chen, C.; He, C.; Pei, H.; Zhang, J. SAR Detection for Small Target Ship Based on Deep Convolutional Neural Network. *Chin. J. Inert. Technol.* 2019, 27, 397–405.
49. Cui, Z.; Zhang, M.; Cao, Z.; Cao, C. Image Data Augmentation for SAR Sensor Via Generative Adversarial Nets. *IEEE Access* 2019, 7, 42255–42268.
50. Guo, Y.; Du, L.; Lyu, G. SAR Target Detection Based on Domain Adaptive Faster R-CNN with Small Training Data size. *Remote Sens.* 2021, 13, 4202.
51. Chang, Y.-L.; Anagaw, A.; Chang, L.; Wang, Y.; Hsiao, C.-Y.; Lee, W.-H. Ship Detection Based on YOLOv2 for SAR Imagery. *Remote Sens.* 2019, 11, 786.
52. Zhang, T.; Zhang, X.; Shi, J.; Wei, S. High-speed ship detection in SAR images by improved Yolov3. In *Proceedings of the 2019 16th International Computer Conference on Wavelet Active Media Technology and Information Processing*, Chengdu, China, 13 December 2019; pp. 149–152.
53. Zhang, T.; Zhang, X. High-Speed Ship Detection in SAR Images Based on A Grid Convolutional Neural Network. *Remote Sens.* 2019, 11, 1206.
54. Shen, F.; Wang, Q.; Jiang, J. Real-Time Target Detection Algorithm Based on Improved Convolutional Neural Network Ship. *Comput. Appl. Res.* 2020, 4, 316–319.
55. Chen, S.; Zhan, R.; Wang, W.; Zhang, J. Learning Slimming SAR Ship Object Detector Through Network Pruning And Knowledge Distillation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 2021, 14, 1267–1282.
56. Zhou, L.; Wei, S.; Cui, Z.; Fang, J.; Yang, X.; Ding, W. Lira-YOLO: A Lightweight Model for Ship Detection in Radar Images. *J. Syst. Eng. Electron.* 2020, 31, 950–956.
57. Zhang, T.; Zhang, X.; Shi, J.; Wei, S. Depthwise Separable Convolution Neural Network for High-Speed SAR Ship Detection. *Remote Sens.* 2019, 11, 2483.
58. Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 21 July 2017; pp. 2117–2125.
59. Zhang, D. Ship Detection of SAR Images Based on Deep Learning. Master's Thesis, Xi'an University of Electronic Science and Technology, Xi'an, China, June 2021.
60. Liu, S.; Kong, W.; Chen, X.; Xu, M.; Yasir, M.; Zhao, L.; Li, J. Multi-Scale Ship Detection Algorithm Based on a Lightweight Neural Network for Spaceborne SAR Images. *Remote Sens.* 2022, 14, 1149.
61. 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.
62. Xu, P.; Li, Q.; Zhang, B.; Wu, F.; Zhao, K.; Du, X.; Yang, C.; Zhong, R. On-Board Real-Time Ship Detection in HISEA-1 SAR Images Based on CFAR And Lightweight Deep Learning. *Remote Sens.* 2021, 13, 1995.
63. Cui, Z.; Quan, H.; Cao, Z.; Xu, S.; Ding, C.; Wu, J. SAR Target CFAR Detection Via GPU Parallel Operation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 2018, 11, 4884–4894.
64. Kang, M.; Leng, X.; Lin, Z.; Ji, K. A modified Faster R-CNN based on CFAR algorithm for SAR ship detection. In *Proceedings of the 2017 International Workshop on Remote Sensing with Intelligent Processing*, Shanghai, China, 18–21 May 2017; pp. 1–4.