# Survey of Research on Sea Battlefield Target Recognition based on Deep Learning

**Lianping Shan [a], Pan Jiang [b], Yihai Liu [c]**

Jiangsu Automation Research Institute of CSIC, Lianyungang 222006, China

[a] 18036672220@189.cn, [b] jsyxjp@163.com, [c] liuyihai@126.com

**Keywords:** Convolutional neural network, deep learning, image recognition.

**Abstract:** In recent years, convolutional neural networks, CNN have become more and more excellent in the fields of image classification and object detection. The research on the application of deep learning in sea battlefield target recognition is more and more abundant. This paper first summarizes the theory and development process of the commonly used deep learning techniques in the target image recognition system. Then, compares traditional recognition technology and deep learning technology, two-stage model based on regional proposal and one-stage model based on regression. The status quo of deep learning technology in the sea battlefield target recognition is reviewed. Finally, we prospect the possible direction of future recognition technology of sea battlefield targets.

## 1. Introduction

In modern sea battlefield, many platforms, which can monitor and capture a large number of images, from where can calculate the target's identity and location information, such as satellites and drones have been widely used in target detection. It is of great value to pre-war planning, decision support, and precision strikes. However, the image quality is usually susceptible, so it is very important to study the advanced automatic target recognition technology.

At present, there are only a few reviews for image target recognition. Chen Wenting et al[1] made a comprehensive summary of the SAR image ship target recognition, but due to the rapid development of deep learning technology in recent years, it is difficult to accurately summarize the current target recognition technology; The research on vehicle target recognition in complex background is reviewed by Xie Xiaozhu et al[2], but the introduction of deep learning is not detailed enough.

The development of the sea battlefield target recognition system has gone through two stages: recognition system based on traditional and deep leaning. This paper will focus on applying of the deep learning, comparing different technologies, and looking forward to the direction of sea battlefield target recognition.

## 2. Sea Battlefield Target Recognition System

### 2.1 Classification Model based on CNN

In 2012, AlexNet [4] won the title of ImageNet competition with absolute superiority. Since then, CNN has made new breakthroughs in computer vision.

CNN usually includes input layer, convolution layer, pooling layer, fully connected layer, and output layer. In convolution layer, a kernel of certain size, performed from left to right, top to bottom at a certain stride, is used to press the output of the upper layer; the pooling layer performs small neighborhood feature point integration on the convolution result; the full connection layer classifies the data after a series of convolution and pooling steps. Back propogation is used to optimize error, all of the parameters in convolution layer and fully connected layer are updated.

In 1998, LeCun et al. [3] proposed the prototype of modern CNN for text image recognition, LeNet, which includes two convolutional layers, two pooling layers, and a fully connected layer, as is shown in Figure 1. The convolutional layer and the pooling layer are used to extract features, then map raw data to feature dimensions, and fully connect layers is responsible for classifying feature dimensions.

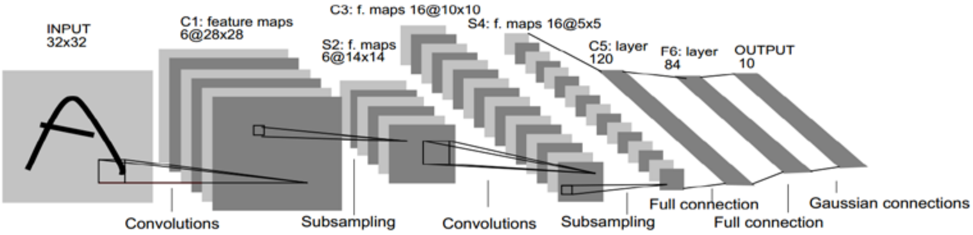

Fig.1 Architecture of LeNet-5, a convolutional network.

AlexNet [4] deepened the layer of the network based on LeNet, using 5 convolutional layers and 3 fully connected layers, which made many improvements, such as: 1) using ReLU to solve disappearance of gradient when the network is deep; 2) using LRN normalization to improve the generalization ability; 3) using dropout, randomly discarding neural in the full connection layer during training and data augmentation, to suppress over-fitting; 4) Distributed training on multi-GPUs. AlexNet made a breakthrough in the ILSVRC2012.

In order to improve the performance of CNN on image classification, researchers at Oxford University proposed a deeper model, named VGG [5], using 3*3 kernels to extract image features in the convolutional layer. In order to fuse multi-scale model features, InceptionNet was proposed [6]. In InceptionNet, to avoid the uncertainty of manually determining kernel size, multiple kernels of different sizes were used to convolve the previous output in the same layer, then concats all of the convolution operations. In addition, except the last layer, all of the layers are used as the feature extractor by reducing the number of fully connected layer.

Deep neural networks are difficult to train because of gradient disappearance and explosion. In response to this problem, ResNet, based on the idea of learning the residuals between input and output, was present by He et al. [7]. In ResNet, each residual block will activate the superimposition of the current layer and shortcut, as is shown in Figure 2.

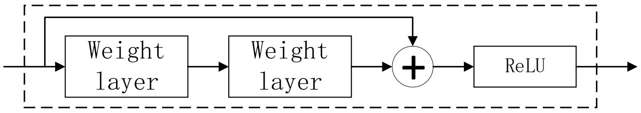

Fig.2 Residual Block

Deep learning, especially CNN, has almost achieved a dominant position in computer vision. Diana et al.[8] proposed an improved CNN to realize remote sensing aircraft recognition; To get a streamlined and easy-training network, Zheng Guangdi et al.[9] optimized the number of layers and nodes of VGGNET, thus, effectively identify the target in the complex sea battlefield environment; deep features, extracted by CNN, edge features, extracted by HOG, and color features, extracted by HSV are combined by Zhao Liang et al[10] to achieve better warship Recognition.

## 2.2 Transfer Learning Techniques

It is difficult to collect a large number of images, which contains real sea battlefield targets, so transfer learning can be considered. By simply adjusting the model trained on a problem, a new model for new problems can be obtained.

Donahue J et al. [11] pointed out that a single-layer fully connected neural network has a good distinction between 1000 classes of images, when the input features are extracted by trained Inception-v3 model. So, the output of last convolutional layer of Inception-v3 model can be used as a streamlined and highly expressive feature vector for any images. Ge W et al. [12] pointed out that freezing the pre-n-layer parameters of the trained CNN model and fine-tuning the rest parameters according to the existing data can be used to obtain a better model. Thus, the more labeled data you have, the smaller the number n is, and the more accurate the model obtained.

### 2.3 Target Detection based on Two-stage Algorithm

It is not accurate to divide the entire image into a single class, because a complete recognition system needs to identify the class of all targets and their specific locations in the image. Like the traditional recognition system, the target detection algorithm based on two-stage algorithm generates proposal regions, and then classify the proposal regions by classification model. In recent years, many mature two-stage algorithms including R-CNN series algorithms have been put forward.

The sliding window detection method violently slides a fixed-size window from left to right, from top to bottom, and classify the region in window by a pre-trained CNN. What the shortcomings of the sliding window detection method is that calculation cost is very high and the size and slides of window is uncertained.

In response to the sliding window detection method, Girshick R et al.[13] proposed a method, named R-CNN, as shown in Figure 3, to extract 2000 regions that may contain the target by selective searching, and then extract the features of these regions by pre-trained CNN, finally excute target classification and border regression.

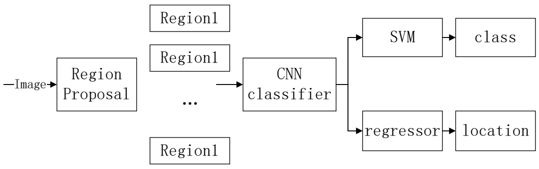

Fig.3 Architecture of R-CNN

Although R-CNN greatly reduces the calculation cost, the 2000 proposal regions must be feed to CNN to do feature extraction, target classification and border regression respectively. To further reduce the calculation cost and solve the repetitive calculation of R-CNN, the feature map, instead of the original image was used to detect target by Girshick R [14], which was named as Fast R-CNN. Fast R-CNN first extracts the entire image feature using CNN, and then uses the proposal region created by the Selective Search method on the feature image, as is shown in Figure 4.

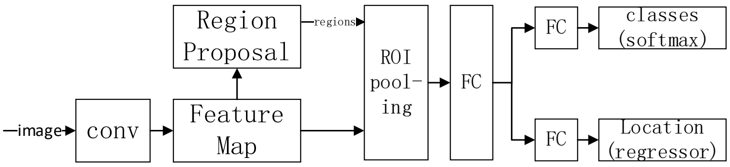

Fig.4 Architecture of Fast R-CNN

For the problem that the 2000 candidate regions are too long to generate for the Fast R-CNN model, Ren S et al. [15] proposed Faster R-CNN, as is shown in Figure 5, which adds a Region Proposal Networks (RPN) after the last convolutional layer. In Faster R-CNN, Prosal regions is generated by RPN and determined whether the candidate region contains a target of a specific category. Finally, the regressor is used to further adjust the proposal region, which includes targets. Faster R-CNN greatly improves the efficiency of target detection and recognition.

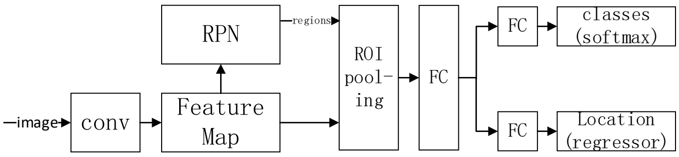

Fig.5 Architecture of Faster R-CNN

## 2.4 Target Detection based on One-stage Algorithm

Faster R-CNN still accomplishes target recognition by generating candidate regions. Many of the candidate regions with large overlaps bring repeated recognition work. Therefore, Faster R-CNN still has high calculational cost.

Redmon J et al. [17] proposed a model, named You Only Look Once (YOLO), which is a one-stage detection model. In YOLO, the target detection is treated as a regression problem. CNN is used to infer the input image only once, and then the location, class and corresponding confidence probability of all objects in the image are directly obtained. There are many another one-stage algorithm, such as YOLOv2[18], YOLOv3[19], SSD [20] and so on, we will not analyze them one by one.

## 3. Comparison

### 3.1 Traditional Technology and Deep Learning Technology

Traditional recognition system divides the process into four steps: preprocessing, feature extraction, feature fusion and recognition, as is shown in Figure 6.

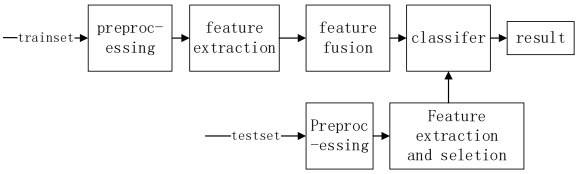

Fig.6 Traditional Sea Battlefield Target Recognition System

Different from the traditional recognition system, the recognition system based on deep learning extracts the important features and finishes target recognition through the automatic learning under a large amount of training data, as is shown in Fig. 7.

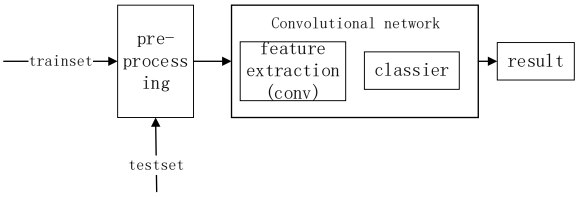

Fig.7 Recognition System Based on Deep Learning Technology

The difference between two type of system is mainly reflected in the following three aspects:

Firstly, the way of feature extraction is different. It is necessary to manually extract a variety of features, and do some feature fusion to remove redundant features in traditional system; CNN attempts to learn features from trainset can greatly reduce the cost of discovering features. Usually, features learned by CNN from a large amount of data are more robust.

Secondly, requirement of the volume and diversity of the trainset is different. As the volume of the data increases, the performance of the traditional recognition system is easily saturated, while the performance of the recognition system based on the deep learning technology can be continuously improved.

Thirdly, the running time of the system is quite different. The training period in traditional classification system is very short, but the feature extraction often involves complex image transformation, so real-time performance of online data prediction is difficult to guarantee. The recognition system based on deep learning has too many parameters to learn during training period, but trained model only involves simple four operations, and will run very fast on GPU duriong prediction period.

### 3.2 One-stage and Two-stage Algorithms

Commonality of both one-stage and two-stage algorithms is that the feature extractors are both CNN, while the way of target detection is different. Two-stage algorithms repurposes classifiers to perform detection. Instead, one-stage algorithms frame object detection as a regression problem to spatially separated bounding boxes and associated class probabilities.

Two-stage algorithms divides the target detection problem into region proposal and target classification. However, one-stage algorithms use a single network and predict bounding boxes and class probabilities directly from full images in one evaluation. Since the whole detection pipeline is a single network, it can be optimized end-to-end directly on detection performance. So, one-stage algorithm is much faster than two-stage algorithm, while the two-stage algorithm is more accurate.

At present, target detection and recognition are still in the research stage, and there are not many practical applications. Hu Yan et al [20] constructed a three-layer convolutional neural network under the framework of Faster R-CNN, and tested the wide-width SAR images of four different ocean clutter environments. Zhou Qi [22] proposed a new YOLO network model by fusing low-level features and abstract features, realizing the real-time detection of mobile ships.

## 4. Conclusions and Prospects

Deep learning is a data-driven technology. At present, because of lack of large scale tagged data in practical applications, the traditional algorithm will still be the main method for sea battlefield target recognition for a long time, but the trend of applying deep learning technology has become more and more obvious.

It is predictable that collecting real tagged data, augmenting data and finding models that can be applied will be the focus of future work. In addition, combining the classical features extracted manually and the abstract features extracted by CNN for classification and using SVM for high-dimensional features to classify features extracted by CNN have been proved to improve recognition accuracy. Finally, the sea battlefield target recognition system has high requirements for the accuracy and efficiency of target detection and recognition. Therefore, the recognition technology based on one-stage algorithms will be the trend of future research.

## Acknowlegement

Defense Pre-Research Fund Project Multi-source heterogeneous information fusion and target recognition technology 3020101020102.

## References

[1] Chen Wenting, Xing Xiangwei, Ji Kefeng. Overview of Ship Target Recognition in SAR Image[J]. MORDEN RADAR, 2012, 34(11): 53-58.

[2] Xie Xiaozhu, He Cheng. Review of Vehicle Target Recognition in Complex Environment Background[J]. Journal of Ordnance Equipment Engineering, 2017, 38(06): 90-94.

[3] Lécun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11):2278-2324.

[4] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]//International Conference on Neural Information Processing Systems. Curran Associates Inc. 2012:1097-1105.

[5] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J]. arXiv preprint arXiv:1409.1556, 2014.

[6] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2015:1-9.

[7] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[8] Chao Anna, Liu Kun. Aircraft target recognition of remote sensing image based on convolutional neural network[J]. Journal of Microcomputers and Applications, 2017, 36(22): 66-69+73.

[9] Zheng Guangdi, Pan Mingbo, Liu Wei, et al. Collaborative identification method of sea battlefield target based on deep convolutional neural network[J]. Optics and Optoelectronic Technology, 2018, 16(02): 20-25.

[10] Zhao Liang, Wang Xiaofeng, Yuan Yitao. Research on Ship Identification Method Based on Deep Convolution Neural Network[J]. Ship Science and Technology, 2016, 38(15): 119-123.

[11] Donahue J, Jia Y, Vinyals O, et al. Decaf: A deep convolutional activation feature for generic visual recognition[C]//International conference on machine learning. 2014: 647-655.

[12] Ge W, Yu Y. Borrowing Treasures from the Wealthy: Deep Transfer Learning through Selective Joint Fine-Tuning[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2017:10-19.

[13] Girshick R, Donahue J, Darrell T, et al. Region-Based Convolutional Networks for Accurate Object Detection and Segmentation[J]. IEEE Trans Pattern Anal Mach Intell, 2016, 38(1):142-158.

[14] Girshick R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.

[15] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Trans Pattern Anal Mach Intell, 2015, 39(6):1137-1149.

[16] Bojarski M, Del Testa D, Dworakowski D, et al. End to end learning for self-driving cars[J]. arXiv preprint arXiv:1604.07316, 2016.

[17] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2016:779-788.

[18] Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger[C]// IEEE Conference on Computer Vision & Pattern Recognition. 2017.

[19] Redmon J, Farhadi A. YOLOv3: An Incremental Improvement[J]. 2018.

[20] Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector[C]// European Conference on Computer Vision. 2016.

[21] Hu Yan, Shan Zili, Gao Feng. Target Detection of Marine Ships Based on Faster-RCNN and Multiresolution SAR[J]. Radio Engineering, 2018, 48(02): 96-100.

[22] Zhou Qi. Multi-target real-time detection of mobile ships based on YOLO algorithm[J]. Computer Knowledge and Technology, 2018, 14(10): 196-197.