Research Article

Sonar Image Target Detection and Recognition Based on Convolution Neural Network

Wu Yanchen

School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, China

Correspondence should be addressed to Wu Yanchen; 202008064@gxstnu.edu.cn

Received 22 February 2021; Revised 10 March 2021; Accepted 14 March 2021; Published 22 March 2021

Academic Editor: Mian Ahmad Jan

Copyright © 2021 Wu Yanchen. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Recent advancements in deep learning offer an effective approach for the study in machine vision using optical images. In this paper, a convolution neural network is used to deal with the target task of sonar detection, and the performance of each neural network model in the sonar image detection and recognition task of underwater box and tire is compared. The simulation results show that the neural network method proposed in this paper is better than the traditional machine learning methods and SSD network models. The average accuracy of the proposed method for sonar image target recognition is 93%, and the detection time of a single image is only 0.3 seconds.

1. Introduction

The targets under the sea are not convenient for the human to contact directly, so sonar is needed for detection and recognition. At present, target detection and recognition of sonar images is an important research topic of marine target detection. Traditional sonar detection mainly identifies the features or contours of marine organisms, such as the shape and texture of underwater targets, but the effect of recognizing these features mainly depends on whether the features are clear. However, due to the complexity of the underwater environment and the similarity of some targets, the traditional sonar recognition cannot meet the current detection and recognition of underwater targets. Today, deep learning target detection and identification becomes the mainstream method, and it can be expressed as depth feature extraction target identification and positioning based on deep convolutional neural network. The convolutional neural network (CNN) model has strong characterization and modeling capabilities through supervision or non-supervision training methods. CNN can be denoted by the characteristics of the object layer, and the abstraction and description of the object hierarchism are achieved. Girshick et al. designed a detector (RCNN) based on regional CNN, which is also a milestone in the field of target detection. Since

then, the target detection is developed quickly, and many RCNN-based methods are developed in the literature.

In recent years, several computational models have been proposed in the literature for the identification and detection of underwater acoustics using machine learning algorithms. For example, Li et al. [1] and Cheng et al. [2] used the deep learning algorithm for data fusion of sonar multisensor observation data. This article sets out a new approach to SAR image object recognition by designing CNN and improves the unattended detection mechanism, taking advantage of the convolution features. Similarly, Yang et al. [3] and Lei et al. [4] proposed another model based on deep learning for the identification of active sonar signal and semantic segmentation of remote sensing images, respectively. Both the authors produce the promising results, and their strategy was based on how to extract the appropriate number of regional recommendations in order to increase the reliability of target detection. Moreover, Zheng [5] and Liu [6] used the method of passive underwater acoustic using deep learning as classification learning algorithm.

To investigate the convolutional neural network approaches used for remote sensing images' semantic segmentation and to create new technologies and algorithms for greater accuracy, in this paper, we proposed the novel computational model based on convolutional neural

network for the sonar image target detection and recognition. The most important additions and the major contributions of the paper are as follows:

- (1) An intelligent and powerful computational model is proposed for the classification of sonar image target detection and recognition
- (2) This work proposes an algorithm that can automatically perform target recognition, tracking, or detection works
- (3) This work proposes a rigorous model that classifies multitype objects at the same time as the traditional approach using a feature matching technique that can detect one type of object at a time
- (4) Finally, the proposed scheme has been extensively tested on comparative experiments and ablation studies

The rest of the paper is organized as follows. In Section 2, literature review is studied in detail, while Section 3 provides the detailed methodology. Section 4 provides detailed results and discussion. Finally, the paper is concluded in Section 5.

2. Design of Convolutional Neural Network Model

In this section, we introduce design of the proposed convolutional neural network model. The architecture of the proposed model is shown in Figure 1 which contains a number of components that are discussed in detail as follows.

2.1. Convolutional Neural Network. A convolutional neural network (CNN) is a kind of deep neural network [8–12]. In the 1990s, Lecun first proposed to design and train convolutional neural network lenet5 by using a backpropagation algorithm, which has achieved good results in handwritten digit recognition [13]. In 2006, Professor Hinton of the University of Toronto in Canada proposed the concept of deep learning and solved the gradient dispersion problem of the traditional neural network through the layer pretraining method, which set off an upsurge of deep learning [14]. At present, the convolutional neural network, as the most widely used deep neural network, has made a breakthrough in image processing [15], speech recognition [16], target detection [4, 5], and other fields. Generally speaking, the basic structure of CNN consists of a convolutional layer, pooling layer, fully connected layer, and output layer, as shown in Figure 1.

2.1.1. Convolution Layer. The convolution layer is composed of multiple feature maps. The convolution layer convolutes the input original image with a certain size and step size through the convolution kernel and obtains the feature map of the next layer after activating the function. Convolution kernel is a weight matrix of size $n \times n$. The convolution process is shown in Figure 2.

Each neuron in the convolution layer is connected with the local region of the upper layer by a set of weights, and the pixel weighted sum is obtained. Then, the locally weighted

FIGURE 1: Structure of convolution neural network, adapted from [7].

FIGURE 2: Convolution operation.

sum is transferred to a nonlinear activation function to obtain the value of each neuron in the convolution layer. The calculation formula is as follows:

$$X_j^i = f \left(\sum_{i \in M_j} X_i^{i-1} * W_{ij}^l + b_j^i \right), \quad (1)$$

where X_j^i represents the j -th feature map of the i -th layer, f represents the activation function, X_i^{i-1} represents the i -th feature map of the upper layer, $*$ represents the convolution operation, W_{ij}^l is the convolution kernel, M_j represents the subset of the feature map of the upper layer participating in the operation, and b_j^i represents the offset. The function of the convolution layer is to extract different features of the input image. In CNN, because the convolution process is linear, to increase the nonlinear expression ability of the network, the nonlinear activation function is usually used. The commonly used activation functions are the sigmoid function, tanh function, and ReLU function. Compared with the other two functions, the ReLU function is widely used because of its faster convergence speed and easier implementation. The ReLU, sigmoid, and tanh activation functions can be mathematically expressed using the following equations:

$$\text{ReLU}(x_i) = \max(0, x_i), \quad (2)$$

$$\text{sigmoid}(x_i) = \frac{1}{(1 + e^{-x_i})}, \quad (3)$$

$$\text{tanh}(x_i) = \frac{2}{(1 + e^{-2x_i})} - 1. \quad (4)$$

2.1.2. Pooling Layer. The pooling layer is also composed of multiple feature maps. The pooling operation is similar to

the convolution operation. Each neuron of the pooling layer is also connected with the local area of the previous layer feature map. However, the value of the pooling layer is not the weighted summation of the local area, but the maximum pixel value or average value of the local area is extracted as the value of the pooling layer neuron. The function of the pooling layer is to reduce the size of the feature graph, reduce the computational complexity of the network model, and improve the spatial invariance of the network for the input image object. The common pooling methods are maximum pooling and average pooling. All the networks in this paper adopt maximum pooling. The pooling operation is shown in Figure 3.

2.1.3. Full Connection Layer. The full connection layer is located behind the convolution layer and pooling layer. Each neuron in the full connection layer is connected with all neurons in the previous layer to integrate the feature extracted from the convolution layer or pooling layer. The output value of the last full connection layer is transferred to the output layer to realize the output of classification results. With the rapid development of the convolutional neural network, the target detection method based on the convolutional neural network has been widely used. Using CNN to extract image target features has gradually replaced the traditional target detection method based on manual features and has become the mainstream method of current target detection. In recent years, many excellent target detection models have emerged among which Fast RCNN, YOLOv3, and SSD models are the most widely used.

2.2. Fast RCNN Model. Fast RCNN is a CNN target detection model based on the region proposal proposed by Hu in 2015 [16]. The basic structure of FastRCNN is still CNN, but this model abandons the traditional selective search algorithm [17] in the process of extracting the possible candidate regions, i.e., region recommendation of the target in the image, and then adds a full convolution network (RPN) after the convolution feature map of the last layer of convolution neural network, and the RPN achieved this function. The overall structure of RCNN is shown in Figure 4.

The input of the model is the original image. The feature map is extracted by a deep convolution neural network, i.e., vgg16 [18], and then the coordinate information of candidate regions is extracted by RPN. About 2000 candidate regions are extracted from each image, and then the coordinate information of these candidate regions is mapped back to the position in the original image and compared with the real region. When the overlap ratio between the candidate region and the target region in the original input image is greater than a certain threshold, the default value is 0.7, the target is considered to exist in the region, and then the regression location is carried out according to the positions of the candidate region and the real region, and the probability distribution of the target belonging to a certain category is calculated. Because Fast RCNN realizes the process of candidate region extraction and target detection

in a neural network, the operation speed of the model is significantly improved compared with RCNN and Fast RCNN.

2.3. YOLOv3 Model. YOLO (you only look once) is a CNN target detection model based on regression thought [11]. YOLO has gone through the improvement of YOLO [19] and YOLO9000 [20], and now it has developed to YOLO3. Different from the Fast RCNN model, the YOLO series model does not need to extract the possible candidate regions of the target in the image but directly performs regression training on the whole image. Firstly, the original image is transformed into a fixed size, and then the feature is extracted by the depth convolution neural network. Finally, the classification result and the target location coordinate information are output.

The basic network used by YOLOv3 is Darknet53 [11]. To avoid the gradient explosion caused by the deepening of network layers, Darknet53 adds RESNET (residual neural network) [21–25] residual structure. When classifying the target objects, YOLOv3 uses several independent logistic regression classifiers. These logistic regression classifiers only judge whether the objects in the target frame belong to the current label or not, which is a simple binary classification. In this way, multilabel classification is realized. Besides, YOLOv3 uses the idea of FPN (feature pyramid networks) [26, 27] for reference and predicts three different scale feature maps, with three target frames predicted for each scale grid. For an image, if it is initially divided into $N \times N$ grids and C categories need to be predicted, then the final predicted tensor is $N \times N \times [3 \times (4 + 1) + C]$, which includes the information of four coordinate points and a confidence score. Due to the fusion of the first two-layer feature maps, the model can obtain more low-level and high-level image semantic information, which makes the model prediction more accurate.

2.4. SSD Model. SSD (single shot multibox detector) is another CNN target detection model based on regression idea and is one of the mainstream target detection models at present. The SSD model is mainly divided into two parts: one is a convolutional neural network for feature extraction, which is based on VGG network, replacing the last two full connection layers of VGG with convolution layer and discarding dropout layer and softmax classification layer; the other is multiscale feature detection network closely following VGG network, which is composed of four convolution layers. Each group first uses 1×1 convolution kernel to reduce the number of channels and then uses 3×3 convolution kernel to increase the number of channels. The feature maps of different levels are used for the border regression of different scale targets and the prediction of different category scores. Finally, the final detection result is obtained by NMS [23]. SSD is combined with a multiscale feature map to detect small targets with shallow feature maps with high resolution and large targets with deep feature maps with low resolution so that targets of different scales can be detected.

FIGURE 3: Pooling operation.

FIGURE 4: Schematic diagram of the FastRCNN model.

3. Model Performance Evaluation Index

To evaluate the performance of a machine learning algorithm, four parameters are commonly used to check the model reliability and validity [10–12]. These parameters

include overall accuracy of the model, specificity as true negative rate, sensitivity as true positive rate, and MCC as Mathew's correlation coefficient. These four metrics calculation formula are as follows:

$$\begin{aligned}
 \text{accuracy} &= 1 - \frac{(m_{-}^{+} + m_{+}^{-})}{(m^{+} + mP^{-})}, \quad 0 \leq \text{Acc} \leq 1, \\
 \text{specificity} &= 1 - \frac{m_{+}^{-}}{m^{-}}, \quad 0 \leq \text{Sp} \leq 1, \\
 \text{sensitivity} &= 1 - \frac{m_{-}^{+}}{m^{+}}, \quad 0 \leq \text{Sn} \leq 1, \\
 \text{Mcc} &= \frac{1 - ((m_{-}^{+} + m_{+}^{-}) / (m^{+} + mm^{-}))}{\sqrt{1 + ((m_{+}^{-} - m_{-}^{+}) / m^{+}) (1 + ((m_{-}^{+} - m_{-}^{-}) / m^{-}))}}, \quad -1 \leq \text{Mcc} \leq 1,
 \end{aligned} \tag{5}$$

FIGURE 5: Example of tire and box sonar.

where

- (i) m^+ represents true positive
- (ii) m^- represents the total true negative
- (iii) m_+^+ represents the false positive
- (iv) m_+^- represent the false negative

4. Results and Discussion

4.1. Data Preparation. In this paper, sonar image data are obtained through experiments. The sonar used in the experiment is m3 multibeam imaging sonar of Kongsberg Company, UK. The working frequency is 500 KHz, the range resolution is 0.01 M, the nearest distance is 0.2 m, and the farthest distance is 150 m. A total of 2300 sonar images were obtained through several experiments, 80% of which were training sets and 20% were test sets. The final sonar image is shown in Figure 5.

The resolution of the sonar image is $1920 * 1080$, the sector area is the scanning range of sonar, the scanning angle is 120, and the origin O of the sector is the scanning center, which is the location of imaging sonar. The longitudinal highlight area on the left and right sides of the image is the pool wall, and the area outside the pool wall is not considered. The white arc is an equidistant line, and the distance from any point on the same equidistant line to O is the same.

4.2. Experimental Setup. In the experiments, we have setup the computing platform with basic hardware and software specifications that include Intel Core i5-10500 processor, NVIDIA Quadro 1060 graphics card, the memory of graphics card is 2G, Tensorflow version is 1.10, CUDA version is 9.0, and operating system is Ubuntu 16.04. To ensure the comparability of the results, the three models use the same super parameters in the training, the learning rate is 0.001, the weight attenuation factor is 0.99, and the momentum parameter is 0.9. The optimization method is random gradient

descent, the number of iterations is 10000, and the number of pictures participating in each iteration is 16.

4.3. Analysis of Target Detection. After the training, the test set is used to test, and the results are shown in Table 1 and Figure 6. According to the results shown in Table 1, in the target detection and recognition task of starfish and scallop sonar images, the map of YOLOv3 reaches 92.95%, and the detection accuracy is much higher than that of the faster can and SSD models. In terms of detection speed, it takes 0.255 s for YOLOv3 to detect an image, which is similar to SSD speed but faster than Fast RCNN which is 4 times faster because Fast RCNN needs to extract about 2000 candidate regions from each image and judge these regions separately, which leads to more time consumption. However, YOLOv3 and SSD models use the regression method to output coordinate information directly, so the detection speed is faster. The example of the detection results of the sonar image of starfish scallop by YOLOv3 shows that the model can still detect the target accurately in the case of large background noise.

Furthermore, the quantity of the sample selected by different algorithm is further direction, and the specific results are shown in Figure 7. Overall, each algorithm can basically maintain the relative stability of data distribution, but the model of training is very different, which further indicates that uncertainty + diversity sampling can not only select high information amount of data but also make the sample have minimum redundancy and high representation.

Through the above experiment, it can be seen that the mode of uncertainty + diversity sampling can play a good effect on the dataset even in the distribution and uneven distribution. Further, the uncertainty + diversity sampling in Figure 7 can obtain a better detection effect than all training data by selecting part of the amount of high information and high representative data in the case where the data are small,

TABLE 1: Detection results of different models on sonar image of starfish scallop.

Test method	Basic network	Picture size	ACC%	Detection speed (s/image)
Faster RCNN	VGG16	1000 × 600	63.03	1.172
YOLOv3	Darknet53	416 × 416	92.95	0.255
SSD	VGG16	300 × 300	31.38	0.224

FIGURE 6: Detection results of the box and tire sonar image by YOLOv3.

FIGURE 7: Continued.

FIGURE 7: Number of samples selected by different algorithms.

and whether the algorithm has a relief the fit of fit that is a direction worthy of more exploration in the future.

5. Conclusions

This paper presented a convolution neural network model for classification and identification of sonar image. The performance of the proposed model was extensively evaluated, and experimental results showed that the proposed YOLOv3 model is effective to predict the box and tire in the sonar image. We also compare the performance of the proposed model with RCNN and SSD algorithms. The results show that the detection accuracy of YOLOv3 is higher than those of the other two methods in the case of tire sonar image target detection task; and the detection speed is almost the same as that of the SSD model, which is four times faster than that of the Fast RCNN model. In the future, we are planning to incorporate and compare, as well as to add tools, the effects of more in-depth learning frameworks such as the RNN and BNN.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The author declares that there are no conflicts of interest regarding the publication of this paper.

References

- [1] G. Li, Li Xiang, X. Tan, and J. Yan, "Target detection in SAR image based on convolutional neural network," *National Security Geophysics Professional Committee of Chinese Geophysical*, vol. 8, 2020.
- [2] L. Cheng, P. Zhou, and J. Han, "High resolution optical remote sensing image target detection based on rotation invariant convolution neural network," *Scientific Observation*, vol. 15, no. 6, pp. 75-76, 2020.
- [3] L. M. Yang, X. Hu, and L. Hong, "Aircraft target detection in remote sensing images based on the deformable convolutional neural network," *Foreign Electronic Measurement Technology*, vol. 39, no. 7, pp. 121-126, 2020.
- [4] X. Lei, *Research on Hybrid Hole Convolution Neural Network for Image Classification and Target Detection*, Xi'an University of science and technology, Xi'an, China, 2020.
- [5] Y. Zheng, *Target Detection and Segmentation of Cervical TCT Image Based on the Convolutional Neural Network*, Xiangtan University, Xiangtan, China, 2020.
- [6] S. Liu, *Remote Sensing Image Target Detection Method Based on Deep Convolution Neural Network*, Guangdong University of Technology, Guangzhou, China, 2020.
- [7] S. Liu and L. Cheng, "Remote sensing image target detection method based on deep convolution neural network," *Industrial Control Computer*, vol. 33, no. 5, pp. 72-74, 2020.
- [8] Q. Cheng, G. Chen, Lu Wang, and C. Guan, "Millimeter-wave image target detection based on the convolutional neural network," *Science Technology and Engineering*, vol. 20, no. 13, pp. 5224-5229, 2020.
- [9] J. Yu, Q. Kang, Z. Chen, M. Chu, Z. Hu, and H. Yao, "Sea surface target detection in remote sensing image based on full convolution neural network," *Journal of Missile and Guidance*, vol. 40, no. 5, pp. 15-19, 2020.
- [10] Y. Huang, *Application of Convolution Neural Network in Cement Electron Microscope Image Target Detection*, Hubei University of Technology, Wuhan, China, 2020.
- [11] Z. Liu and Y. Li, "The target detection method of SAR image based on feature fusion convolutional neural network," *Microprocessor*, vol. 41, no. 2, pp. 31-37, 2020.
- [12] S. Yang, *Remote Sensing Image Target Detection Based on Deep Convolution Neural Network*, Hebei Normal University, Shijiazhuang, China, 2020.

- [13] X. Chen, *Research on Optical Remote Sensing Image Target Detection Technology Based on Convolution Neural Network*, Harbin Institute of Technology, Harbin, China, 2020.
- [14] P. Yin, *Research on Aerial Image Target Detection Algorithm Based on Full Convolution Neural Network*, Harbin Institute of Technology, Harbin, China, 2020.
- [15] G. Chen, Q. Cheng, C. Huang, Z. Zhou, and L. Wang, "Millimeter-wave image target detection using deep convolution neural network method," *Telecommunication Technology*, vol. 59, no. 10, pp. 1121–1126, 2019.
- [16] Q. Hu, "Remote sensing image target recognition and detection based on deep convolution neural network," *Journal of Ningxia Normal University*, vol. 40, no. 10, pp. 75–79, 2019.
- [17] Q. Yao, X. Hu, and L. Hong, "Aircraft target detection in remote sensing images based on multi-scale fusion feature convolution neural network," *Chinese Journal of Surveying and Mapping*, vol. 48, no. 10, pp. 1266–1274, 2019.
- [18] H. Jin, *Remote Sensing Image Target Detection Based on SSD Convolutional Neural Network*, Hebei University of Geosciences, Shijiazhuang, China, 2019.
- [19] W. Zeng and J. Lu, "Sonar image target detection and recognition based on convolution neural network [a]," in *Proceedings of the 17th symposium on ship underwater noise*, vol. 7, Key Laboratory of Equipment Comprehensive Support Technology of University of National Defense Science and Technology Chinese Society of Shipbuilding Engineering, Beijing, China, January 2019.
- [20] H. Li, C. Li, A. Dubai, and J. Ren, "Remote sensing image target detection based on an improved convolutional neural network with an attention mechanism," *Chinese Journal of Image Graphics*, vol. 24, no. 8, pp. 1400–1408, 2019.
- [21] X. Wang, Z. Guan, J. Wang, and Y. Wang, "Color image sonar target detection based on convolution neural network," *Computer Applications*, vol. 39, no. S1, pp. 187–191, 2019.
- [22] C. Hu, C. Chen, C. He, H. Pei, and J. Zhang, "Ship small target detection in SAR image based on deep convolution neural network," *Chinese Journal of Inertial Technology*, vol. 27, no. 3, pp. 397–405, 2019.
- [23] P. Zhang, *Research on Multi-Scale Remote Sensing Image Target Detection Method Based on Convolution Neural Network*, Xi'an University of Electronic Science and Technology, Xi'an, China, 2019.
- [24] Q. Wang, *SAR Image Denoising and Target Detection and Recognition Based on the Convolutional Neural Network*, Xi'an University of Electronic Science and Technology, Xi'an, China, 2019.
- [25] F. Guo, *Research on Remote Sensing Image Target Detection Algorithm Based on Deep Convolution Neural Network*, Harbin University of science and technology, Harbin, China, 2019.
- [26] T. Zhou, *Image Classification and Target Detection Based on the Convolutional Neural Network*, Guangxi Normal University, Guilin, China, 2019.
- [27] L. Fang, *Research on High-Resolution Image Target Detection Based on Deep Convolution Neural Network*, Liaoning University of Technology, Jinzhou, China, 2019.