

Research Article

Sonar Image Target Detection Based on Deep Learning

Shibing Yu 

School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200072, China

Correspondence should be addressed to Shibing Yu; yushibing@baosight.com

Received 2 December 2021; Revised 4 January 2022; Accepted 13 January 2022; Published 18 February 2022

Academic Editor: Gengxin Sun

Copyright © 2022 Shibing Yu. 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.

Due to the nonlinearity of the imaging of sonar equipment and the complexity of the underwater sound field environment, the gray level of the target area of the acquired underwater sonar image is relatively small. These characteristics are the target of the subsequent sonar image. Work such as detection and location tracking has brought great challenges. It has brought great challenges to solving the work of positioning and tracking, which makes the research of sonar image target detection based on deep learning very important. This article aims at studying the use of sonar to detect image targets based on deep learning technology. This article proposes a variety of sound image denoising methods based on multiresolution tools. The purpose of this article is to divide the natural image into blocks at an appropriate rate according to the change of the sampling matrix and measure the underwater natural image. The sound image defines an information model. These methods have greatly changed the image and period of using remote and temporary information. The translation results of these methods are all valid. The sharpening separation method based on filtered image and bidirectional detection should be published through a solution algorithm and different frames, and the expected algorithm can be reused and extracted as an action to improve the similarity of the image and should be saved and separated in detail. The result is correct. This article studies the application of deep learning methods in sonar image target detection and designs corresponding algorithms for improvement and functional realization in view of the current deficiencies and needs in this field. The experimental results show that the improved scheme and applied algorithm proposed in this article can achieve good results, the verification sample set includes 184 remote-sensing aircraft targets, and the resolution of remote-sensing images is unified to 1644×971 size. The accuracy of the target detection algorithm has been significantly improved, reaching 74.6%, and the detection speed has also been greatly improved. Compared with the RNN algorithm, the speed has been increased by 7 times. The detection results confirmed that the improved algorithm has higher positioning accuracy and faster detection speed.

1. Introduction

1.1. Background. My country's sea area is large, covering an area of about 3 million square kilometers. Due to historical reasons, the 3 million sea areas related to China have many disputes in the field of marine environment and maritime rights; however, maritime security is also closely related to China's territorial security. However, there are two traditional methods for determining and determining shipping goals: incentive and noninspiration; the other inactivity is to use the physical field information related to the sound field generated around the ship to analyze the ship, but these traditional methods of ship target recognition are not applicable. This study proposes a new global learning algorithm to recognize ship images [1]. At present, many traditional target detection methods perform well for a certain type of

fixed target or a detection task in a specific scene, but they are not suitable for multiple target detection in a complex environment. At the same time, detection speed is also an important indicator. In many applications, real-time performance is required for target detection, which cannot be met by traditional target detection algorithms. Deep learning, especially the excellent performance of the convolutional neural network in image recognition, makes the target detection and recognition based on deep learning a research hotspot in recent years.

1.2. Significance. The specific image of the sonar is composed of three parts: the shadow area, the seabed review area, and the high target area. In order to process some image content and then move the target partition, and reach the

shadow area from the complex bottom-up inspection of the seabed, eliminate the separation of the target and the background, and minimize the distortion, the accurate specific image distribution is the prerequisite for extracting the main underwater target elements, and the parameter measurement of the underwater target and analysis and identification. The purpose of target recognition, positioning, and tracking is to determine the main characteristics of the target by assigning the moving target and the background, and compare the system capability images of multiple targets to match the target with high similarity. Since the nonlinear image is composed of specific equipment and has a complex underwater acoustic environment, the characteristic of a specific underwater image is that the specific image has a higher gray level in the background noise. The gray level of the target area in the image is relatively small [2, 3]. These target characteristics are very difficult. Questioning the limitations of signal processing is a high-tech subject of signal processing. These contradictions can be resolved in the fields of image design, image separation, image coordination, and image recognition. In addition, the huge amount of calculation and huge model file storage occupies a serious obstacle to the popularization and application of deep learning target detection. Therefore, optimizing the deep learning target detection neural network, reducing the amount of parameters, and reducing the storage space occupation on the premise of ensuring that the detection accuracy is not significantly reduced have important research value.

1.3. Related Work. Marine ship monitoring has always been a research hotspot in countries all over the world, but due to the difficulty of obtaining sea surface images and more interference, the research on ship target detection has certain limitations. Zhu et al. pointed out that the latest advances in deep learning provide an effective method for the use of optical images for machine vision research. This study uses convolutional neural network to deal with the task of sonar detection and compares the performance of each neural network model in the task of underwater box and tire sonar image detection and recognition. The simulation results show that the neural network method proposed in this article is better than traditional machine learning methods and Single-Shot MultiBox Detector (SSD) network models. The average accuracy of the proposed sonar image target recognition method is 93%, and the detection time of a single image is only 0.3 seconds. But only in a limited number of application scenarios, deep learning algorithms cannot make unbiased estimations of the laws of the data [4]. In order to achieve good accuracy, a large amount of data is needed. This requires a lot of manpower and material resources [5]. Tueller et al. pointed out that powerful target detection in sonar images is an important task for underwater exploration, navigation, and mapping. Current methods make assumptions about the shape, highlights, or shadows of objects, which may be invalid for certain environments or targets. We focus on the detection field based on feature extraction. The field does not rely on information about

object shapes but is a robust framework aimed at object detection for a variety of underwater world structures and object types. The proposed framework first estimates the seabed type according to the spatial distribution of the features to determine the best parameter set and then obtains a set of features, which are filtered according to the intensity and distribution to produce detection decisions. The proposed method also provides a method to determine the type of seabed and a method based on machine learning to select the parameters of the feature detector to match the type of seabed evaluated, but the accuracy of the matching evaluation result is not particularly high [6]. Lubis et al. proposed an image denoising method based on two-dimensional wavelet transform and applied it to the seabed recognition data acquisition system. Two-dimensional Haar wavelet proposes a unified framework in image processing, which is used for wavelet image compression and combined with side scan sonar image. There are 7 target detections in the side scan sonar imaging results of the seabed recognition target. Analyze the vibration signal to facilitate fault diagnosis. The signal obtained is the time-domain signal. The experimental results show that the application of the two-dimensional wavelet transform image denoising algorithm can achieve better subjective and objective image quality, which is helpful to collect high-quality data and analyze the image of the data center, which has the best time-domain signal characteristics [7–9].

1.4. Main Content. After comprehensive analysis and analysis of the above transformation methods and educational images, an acoustic image mechanism was established to help illustrate the consequences of water. Without comparison, it is difficult to distinguish garbage and details. Then, three acoustic images based on multiple solutions are provided. In the simulation research, the quality and effectiveness of these experimental methods are evaluated and analyzed. The classic image segmentation method [10] is evaluated and analyzed. The high-resolution subsequent images are very random. The target area is not distance. Train the acoustic image segmentation method to detect and monitor the moving target in the changing sequence. The monitoring performance of the algorithm is analyzed and compared with the change mode. Many cameras are trained to use them. The high resolution is rich in distribution and noise sharing in the channel sequence, and it is difficult to distinguish distance and detail from poverty.

2. Signal Detection Experiment Based on Deep Learning

2.1. Preprocessing of Sonar Images. The imaging principle of sonar images is different from ordinary optical images, and the imaging environment is complex and changeable, often accompanied by a lot of non-Gaussian noise. Therefore, traditional target detection methods based on the assumption of Gaussian noise cannot be applied to sonar images. The image used before is not only the core of image segmentation, but also the low-level image recognition [11]. The

actual image used contains more and more technologies. The following impact on images is to reduce the impact of different sounds and to target different types of images. Therefore, we must find a good preprocessing algorithm to improve the authenticity of the segmentation. In this article, the last image is processed in two steps: one is the pressure processing of the image [12], and the other is the normalization processing of the image. This indicates that the image should be used first. Then, the image will reflect the subsequent image segmentation and the degree of change of the background image.

The performance of the system depends on the sound of the background plan. In order to achieve long-term display, the difference between the signal field and the sound field should be fully utilized to achieve the visual indication function. It is very important to understand the statistical characteristics of the background sound field. The main source of ocean sound is the result of very random factors. According to the central theory, ocean sound is an influential signal with a Gaussian distribution to a certain extent, and its echo is a special interface form in the rest wave process. This is because there are many ignorant distributors in the sea [12], which eventually exploded the distributor signal from the signal end. This is part of the background sound, which limits discovery at close range. The amplitude around it follows the Rayleigh distribution and follows the uniform distribution of certain characteristics. The high-resolution image follows the high operating frequency and low light, and has relatively good antireverberation ability, but the sharp lens works late, so changing the level is very uncomfortable. Then, due to the temperature and safety of the ocean, the volume of the sound speed changes to calculate the volume, which changes the volume of the sound wave, and thus the volume of the sound wave. In fact, sound waves are linear. High-gravity images are affected by fish and marine life (such as water and marine life), so it produces undesirable light spots. This is an influential message to choose and define which can be seen as an influential voice. According to the above analysis, before the subsequent image segmentation and definition classification, in order to maximize the retention of attribute information and interface information on the image, these sounds, averaged in the filtered and frequency domains, must be transformed into adaptive soft filtering and waveforms in different ways [13].

2.2. Traditional Target Detection Algorithm. The early multilayer perceptrons had fewer layers, and the connection structure was relatively single, and most of them were fully connected layers. The deep learning neural network is a multilayer perceptron with multiple hidden layers. The deep learning neural network can continuously combine and process abstract low-level features to extract high-level features, which greatly improves the learning ability of the network.

Traditional target detection algorithms include deformable part model, which divides the object into multiple parts and obtains a part template for each part [4].

Traditional target detection algorithms generally use a multiscale sliding window frame [14]. The flowchart of the traditional target detection framework contains the main key steps: target location, feature extraction, and classifier classification. Select the application area from the image through sliding windows of different sizes to achieve the visual effect of the application area; use the classified display to optimize the window. Commonly used classifiers are support vector machine (SVM), Adaboost, and so on. SVM obtains the support vector of the classification plane by maximizing the classification interval and has good classification accuracy on small linearly separable datasets. In addition, by introducing the kernel function to map the low dimensional to the high dimensional, it is widely used in the detection scene.

2.2.1. Target Positioning. The position of the target in the image is uncertain, and the size and shape of the target are also uncertain. Therefore, the multiscale sliding window operation is used to exhaust image blocks of different scales and different aspect ratios [12, 15]. Detect and frame the target location. Multiscale sliding window operation is essentially an exhaustive method. The multiscale sliding window contains all possible regions of the target to be tested, which greatly reduces the missed detection rate, but at the same time, it also makes the traversal time too long and generates too many candidate regions, which changes the time complexity of subsequent feature extraction and classifier classification. The complexity of this method is too high to meet the real-time needs of the detection field. In addition to the typical multiscale sliding window method, target localization methods also include regular block, selective search, and other methods. In view of different situations, the choice of target location algorithm is very flexible.

2.2.2. Feature Extraction. Various types of targets have their own characteristics. To find out the common characteristics between similar targets and different characteristics between different types of targets, it is necessary to manually design a feature with high robustness [16]. Therefore, the quality of feature design and selection directly affects the accuracy of subsequent detection. This article takes SIFT feature extraction as an example to introduce the process of feature extraction. The following details the process steps of the SIFT algorithm.

First, use Gaussian blur to obtain the scale space of the image $L(x, y, \delta)$:

$$L(x, y, \delta) = G(x, y, \delta) * I(x, y). \quad (1)$$

The image is $I(x, y)$, and the two-dimensional Gaussian function [17] is set to $G(x, y, \delta)$. The original image is downsampled, and Gaussian filtering is performed to build an image Gaussian pyramid. Subtract each group of two adjacent layers of images in the image Gaussian pyramid to obtain a Gaussian difference image, and traverse all scales to form a Gaussian difference pyramid:

$$\begin{aligned} D(x, y, \partial) &= (G(x, y, k \partial) - G(x, y, \partial)) * I(x, y) \\ &= L(x, y, k \partial) - L(x, y, \partial). \end{aligned} \quad (2)$$

The difference of Gaussian operator is used to detect the extreme value, and the local features of the image are obtained. After obtaining the Gaussian difference pyramid, determine an extreme point by comparing the size of the pixels in the same layer and the upper and lower neighborhoods of the pixels. The extreme points of the discrete space may have jumps, disturbances, etc., so it is not necessarily a local area. Therefore, the extreme point is accurately located by straight line fitting, so the extreme points are accurately located by straight-line fitting, and the fitting function is

$$D(x) = D + \frac{\ell D^T}{\ell X} X + \frac{1}{2} X^T \frac{\ell^2 D}{\ell X^2} X. \quad (3)$$

Here, $X = (x, y, \partial)^T$; take the above formula to derive the extreme point where the derivative is zero, and the offset of the extreme point is

$$\hat{X} = \frac{\ell^2 D^{-1} \ell D}{\ell X^2 \ell X}. \quad (4)$$

The extreme value corresponding to the extreme point is

$$D(\hat{X}) = D + \frac{1}{2} \frac{\ell D^{-1}}{\ell X} \hat{X}. \quad (5)$$

After the extreme points are obtained by fitting the curve [18] by the above method, the extreme points still need to be optimized to remove unstable edge points and low-contrast points.

For the determined key point position, determine the direction of the key point, and the reference direction of the key point is determined by the neighborhood where the key point is located. Draw the histogram of the pixel gradient direction in the neighborhood of the feature point. The peak of the gradient direction histogram represents the feature point and the gradient direction of the neighborhood, and the maximum value represents the main direction of the feature point. The modulus $m(x, y)$ of the gradient vector and the direction $\theta(x, y)$ of the gradient are as follows:

$$\begin{aligned} m(x, y) &= \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \\ \theta(x, y) &= \arctan\left(\frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)}\right). \end{aligned} \quad (6)$$

2.3. Basic Theory of Deep Learning Algorithms. At present, the convolutional neural network has the most excellent performance in the target detection algorithm and image recognition. This article uses the convolutional neural network to improve the detection algorithm.

2.3.1. Principle of Convolutional Neural Network [19]. Convolutional neural network is composed of a convolutional layer and a pooling layer [20, 21]. Pooling operation is a special convolution process. Convolution and pooling simplify the model complexity. The following briefly introduces the principle of the convolutional neural network: the constituent unit of neural network is called neuron, and the corresponding formula is as follows:

$$h_{w,b}(x) = f(W^T X) = f\left(\sum_{i=1}^3 W_i x_i + b\right). \quad (7)$$

Combining multiple neurons is a neural network model, as shown in Figure 1.

Use the back propagation algorithm to train the convolutional neural network. First, calculate the output a_i of the hidden layer node [22] and the output y_i of the output layer node in the network, and then calculate the error term ε_i of each node. For the output layer node i ,

$$\varepsilon_i = y_i(1 - y_i)(t_i - y_i). \quad (8)$$

Here, ε_i is the error term of node i , y_i is the output value of node i , and t_i is the target value of the sample corresponding to node i . For hidden layer nodes,

$$\varepsilon_i = a_i(1 - a_i) \sum_{k \text{ outputs}} w_{ki} \varepsilon_k. \quad (9)$$

Here, a_i is the output value of node i , and w_{ki} is the node connection weight. Finally, update the weight on each connection:

$$w_{ji} \leftarrow w_{ji} + \eta \varepsilon_j x_{ji}. \quad (10)$$

Here, w_{ji} is the weight from node i to node j , η is the learning rate constant, and x_{ji} is the input passed by the node to node j .

Confidence reflects whether the current bounding box contains the target and the accuracy of the target position. The calculation method is as follows:

$$\text{confidence} = P(\text{object}) * IOU. \quad (11)$$

3. Signal Detection Experiment Based on Deep Learning

3.1. Experimental Description and Conditions. The purpose of the experiment is to prove the effectiveness of the proposed method by using the same feature map to predict the two tasks of location prediction and category prediction. In

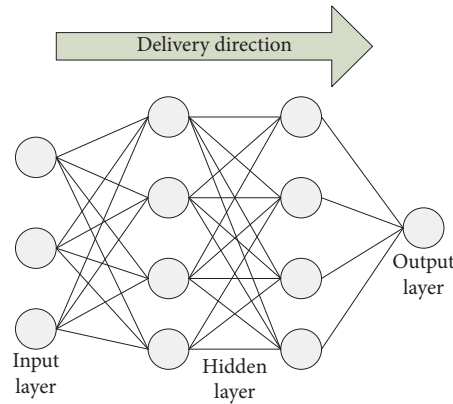


FIGURE 1: Schematic diagram of the traditional neural network structure.

this section, we will analyze the performance of the GC signal design method proposed in this article [23, 24] and the corresponding initial phase optimization search method. In the simulation, we will change part of the experimental conditions, including signal-to-noise ratio, signal mixing ratio, and the radial velocity of the target to verify the detection effect of the signal. The simulation of the primary phase optimization method mainly compares the calculated results with the PAPR values obtained by the two commonly used optimization methods and simulates the number of different line spectra to verify the effectiveness of this method. The experimental condition is to consider a single-transmit and single-receive active sonar system. The signal is first narrow-band-filtered at the receiving end. The bandwidth of the prefilter is (800, 2200) Hz, and then, the multicopy correlation method is used to obtain the signal detection and velocity estimation, the system sampling rate is 10 kHz, and the sound velocity in water is set to 1500 m/s; regardless of the sound ray bending and the unevenness of the sound velocity in the water, the direction of the target away from the sound source is the positive direction of the velocity, and the echo signal is simulated by interpolation obtained legally.

Example 1. According to the idea of this article, design a two-frame GC signal, the starting frequency of the signal is 1 kHz, the total bandwidth of the signal is about 1 kHz, and the signal pulse width is 100 ms, and the target maximum radial velocity is 20 m/s. The frequency points are calculated at 8 m/s, and the antispeed ambiguity frequency interval df is 2 Hz and 1 Hz. The two frames of signals are denoted as SVH and SVL, and the numbers of frequency points are 19 and 29, respectively. The frequency spectrum of the signal is shown in Figure 2.

The distance and velocity ambiguity diagrams of the two frames of signals [25] are shown in Figure 3.

According to Figure 3, the time delay resolution capability of the two frames of signals [26] is first analyzed. Assuming that the radial velocity of the target is 6 m/s, the echo signal appears at 10 ms, and the signal-to-noise ratio is -5 dB and -15 dB. The matched filter output of the frame signal is shown in Figure 4.

It can be seen from Figure 4 that when the signal-to-noise ratio is high, both signals can get better time delay

peaks, but due to the different signal spectral widths, the amplitude of the side lobes is different. When the signal-to-noise ratio decreases, up to -15 dB, the SVL signal can still guarantee a high delay peak output, and although the signal SVH can also have a maximum value at the target appearing position, the larger side lobe has seriously affected the number of echoes.

Example 2. The starting frequency of the signal is 1 kHz, the number of frequency points is different, the initial phase of the signal is optimized by the Newman method [11] and the Narahashi method, and the PAPR is obtained, and the Newman fast algorithm is actually an agglomeration algorithm based on the idea of greedy algorithm. It can also be regarded as a generalized clustering algorithm that can be used to analyze complex networks with 1 million nodes. And then the phases obtained by the two methods are used in this article. The search method is optimized, the search step is $d\phi = \pi/200$, and the optimized phase is obtained and the corresponding PAPR value is obtained. Calculate the value of each frequency point according to the maximum speed tolerance of $v = 10$ m/s, and the comparison result obtained is shown in Table 1.

The frequency points are calculated according to the maximum speed tolerance of $v = 25$ m/s, and the comparison results obtained are shown in Table 2. It can be seen from the comparison of the two graphs that as the number of spectral peaks increases, the peak-to-average power ratio of the signal tends to a relatively stable value. Under the same number of spectral peaks, the signal with a larger spectral peak spacing has larger PAPR, and the results calculated by the Newman method and Narahashi method are not much different. Through comparative analysis, it can be seen that the optimization method in this article can effectively reduce the PAPR of the signal [27], and the average reduction is about 1~2 dB.

Next, we analyze the improvement of the calculation speed of the optimization method and use the number of searches as the criterion. The search step is $d\phi = \pi/200$. Compare the results of the method in this article with the number of searches required for direct phase search. The results obtained are shown in Table 3.

As can be seen from Table 3, as the number of frequency points increases, the calculation amount of the direct search

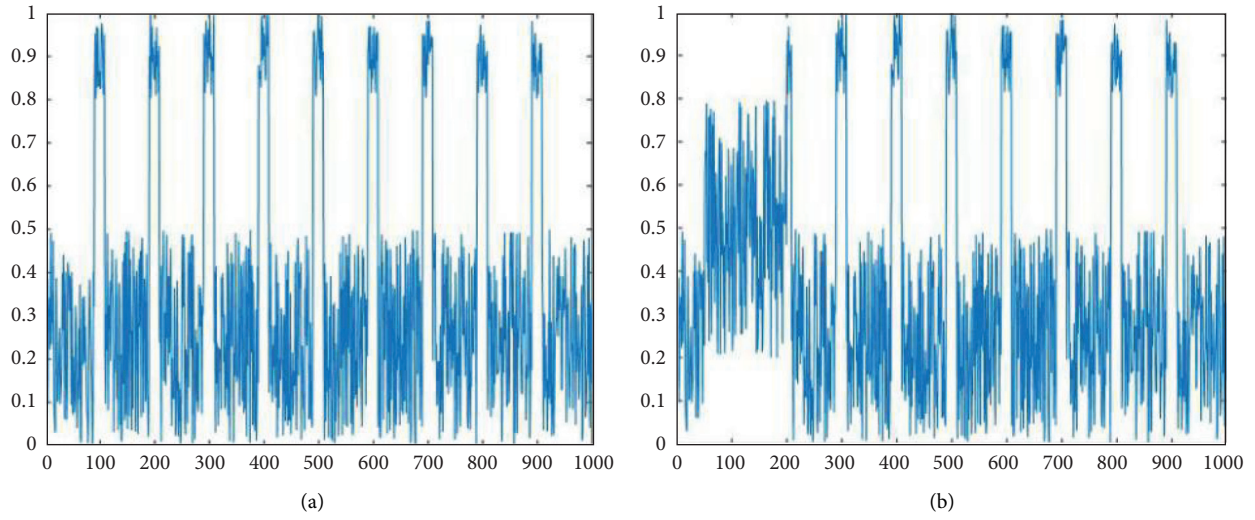


FIGURE 2: Two-frame signal spectrum. (a) SVH spectrum. (b) Spectrum of SVL.

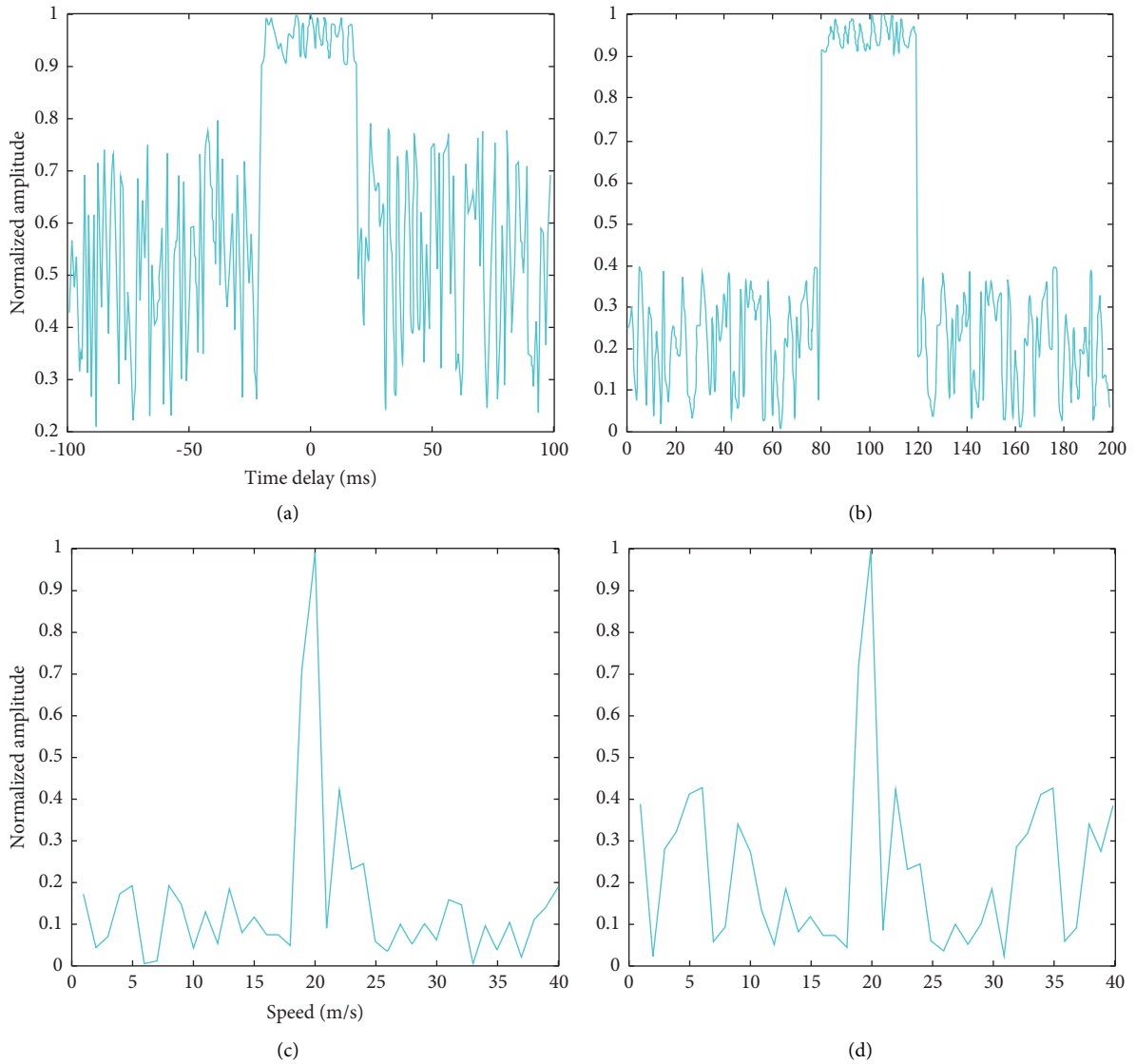


FIGURE 3: Distance and velocity ambiguity diagram of two frames of signal. (a) SVH time delay ambiguity graph. (b) Delay ambiguity diagram of SVL. (c) SVH speed ambiguity graph. (d) Speed ambiguity graph of SVL.

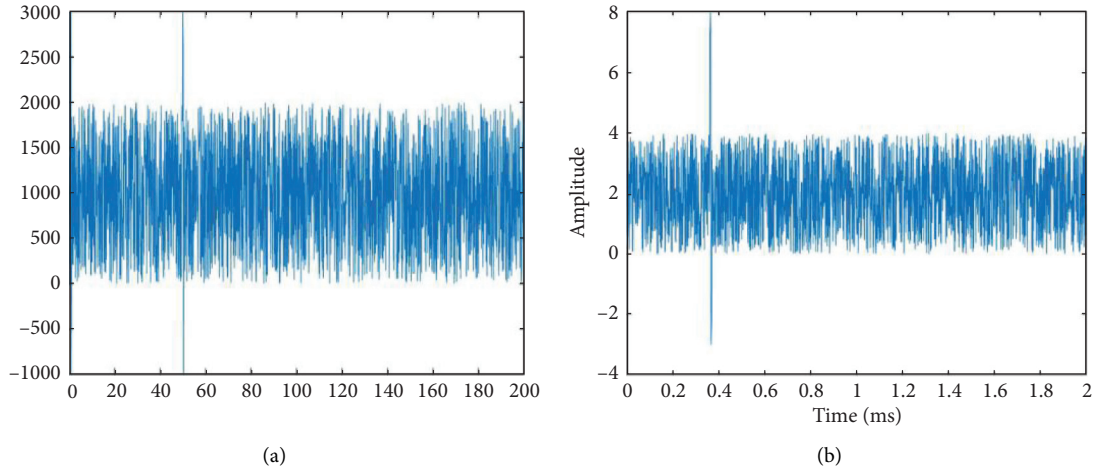


FIGURE 4: Matched filtering results under different signal-to-noise ratios. (a) SVH, SNR = -5 dB. (b) SVL, SNR = -5 dB.

TABLE 1: Year-end assessment form.

Number of frequency points	Newman method	Optimize search method		
		PAPR/dB		
8	6.53	6.20	7.20	5.99
10	7.40	6.44	7.91	6.9
12	8.21	6.47	8.09	6.65
20	7.50	6.96	7.35	6.77
16	8.37	6.98	8.56	7.05
24	7.49	6.81	7.57	7.19
28	8.20	7.25	9.15	7.36
32	9.20	7.84	8.28	7.20
36	7.47	6.93	8.20	7.17
40	8.88	7.77	8.84	7.80

TABLE 2: Comparison of PAPR results of different methods when $v = 25$ m/s.

Number of frequency points	Newman method	Optimize search method		
		PAPR/dB		
8	8.65	7.63	8.59	7.92
10	8.37	7.15	8.58	7.53
12	8.68	7.88	9.09	7.56
16	9.01	7.64	9.04	8.01
20	9.49	8.24	9.97	7.83
24	10.06	8.39	9.68	7.73
28	10.49	8.55	10.69	8.88
32	9.69	7.97	10.22	8.06
36	9.32	8.66	9.71	8.64
40	9.12	8.05	10.35	8.51

TABLE 3: Comparison of search volume between optimization methods and direct methods.

Number of frequency points	8	15	20	25	30	35	40
Optimization	3200	7000	8000	1000	12000	14000	16000
Direct calculation	400^8	400^{15}	400^{20}	400^{25}	400^{30}	400^{35}	400^{40}

method increases exponentially. When the number of frequency points is large, the direct method requires a very large number of searches, which is almost difficult to achieve in real time.

4. Comparative Analysis of Results

4.1. Performance Analysis of Comb Spectrum Signal. The comb spectrum signal [28] has better anti-low Doppler and high Doppler performance than traditional signals. At the same time, this signal combines the long CW signal speed measurement ability and the LFM signal delay resolution ability. With good advantages, better comprehensive resolution of time delay and frequency shift can be obtained. The ambiguity function graph is discretely distributed along the Doppler axis, so that under low Doppler conditions, the overlap area with the reverberation must be less than that of the same length. The CW signal, because of its single sublobe, has narrow-band signal characteristics, and under high Doppler conditions, it can be separated from the reverberation band like a long CW signal and thus suffer less reverberation interference.

4.2. Improved Comb Spectrum Signal Design. From the above analysis of typical comb spectrum signals, it can be seen that GC signals not only have good resistance to low Doppler [29] and high Doppler reverberation, but also have better time delay resolution capabilities. The key to good detection capability of the GC signal is to increase the number of spectral peaks as much as possible without producing velocity ambiguity. If the frequency point interval is too small, Doppler velocity measurement ambiguity will occur, and too large interval means as the number of spectral lines within the limited bandwidth decreases, this will affect its delay estimation capability. When the relative velocity between the sonar and the target is v , in order to prevent Doppler aliasing from the adjacent frequency points of the GC signal, the adjacent spectral peaks should satisfy the following equation:

$$f_i - f_{i-1} > \xi_i = \frac{2f_i v}{c + v}. \quad (12)$$

Simplification can get

$$f_i \left(1 - \frac{2v}{c + v} \right) > f_{i-1}. \quad (13)$$

According to this method, the commonly used GC signal frequency calculation formula is obtained as follows:

$$f_i = f_{i-1} \left(1 + \frac{2v_{\max}}{c - v_{\max}} \right). \quad (14)$$

It can be seen from the above derivation that if the frequency points are designed according to the traditional method, when the velocity is v , the spectrum of the echo signal will move in the opposite direction to the target and the echo spectrum when the velocity is $V_{\max} - v$. There is only one spectral line that does not coincide. At this time, Doppler velocity measurement will be blurred. The signal velocity interval is 30 m/s, and the target radial velocity is

10 m/s and -20 m/s. The echo signal spectrum is divided by one spectrum. In order to prevent this situation, the maximum speed tolerance of the system is usually increased. If the speed detection range of the system is $-30 \sim 30$ m/s, $V_{\max} = 60$ m/s should be used when calculating the frequency. Although this can ensure that the speed measurement ambiguity does not occur, it causes a great waste of frequency bands, which is almost intolerable under the condition of an underwater acoustic channel where frequency band resources are already very limited.

4.3. Comparative Analysis of Experimental Results. In order to further evaluate the effect of the detection algorithm in this article, this section respectively analyzes the detection algorithm based on the bag-of-words model and the Hough location detection algorithm based on the bag-of-words model in the third chapter. All experiments in this article are carried out in the abovementioned hardware and software experimental environment, using the same remote-sensing image aircraft target sample database, and setting the training and verification sample sets according to the ratio of 8 : 2. The verification sample set includes 184 remote-sensing aircraft targets, and the resolution of remote-sensing images is unified to 1644×971 . The training process is given as follows: extract the network in the training area to generate candidate areas. Input the obtained candidate area frame into the position and frame regression network. Initialize the weights of the region extraction network using the parameters learned by the position and border regression network, and the weights of the shared convolutional layer remain unchanged, only minor adjustments are made to the weights of the layers included in the region extraction network, and the position and border regression are fine-tuned. The fully connected layer in the network makes the two networks share the same convolutional layer to form a unified network. During the training process, observe the change trend of the loss value. When the loss value is small enough and tends to be stable, the network model is considered to converge and an effective target detection model is obtained.

The SSD target detection algorithm is a detector based on a fully convolutional network [30], using the features extracted from multiple different layers to directly detect areas of several sizes, which effectively reduces the amount of calculation and further accelerates the speed of target detection. It is characterized by an end-to-end detection process and uses global characteristics, which use different layers to detect objects of different sizes. There is a contradiction in the improvement of the overall accuracy of target detection. The previous feature layer image has a large spatial resolution, but lacks the semantic information of the latter several layers. There is more semantic information in the latter several layers, but after too many pooling sampling operations, the space resolution is too small. To detect small objects, not only a large enough spatial resolution is needed to provide finer features and denser sampling, but also enough semantic information is needed to distinguish it from the background. Small targets tend to rely more on

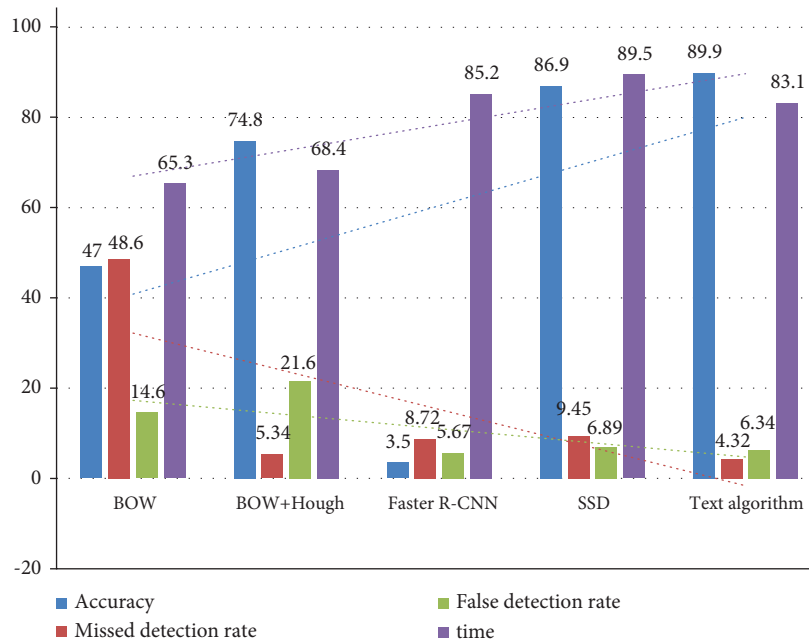


FIGURE 5: Histogram of comparison of evaluation index results of target detection algorithm.

shallow features, because shallow features have higher resolution. Therefore, the SSD target detection algorithm leads to poor detection of small targets.

As can be seen from Figure 5, the Hough voting positioning target detection algorithm based on the bag-of-words model [28] and the Hough voting positioning target detection algorithm based on the bag-of-words model have similar computing times, and the improved Hough voting positioning target detection algorithm based on the bag-of-words model. The accuracy rate has been significantly improved, reaching 74.6%. The fast r-cnn, SSD, and target detection algorithms studied in this article are all target detection algorithms based on convolutional neural networks. These three algorithms are far ahead of traditional target detection algorithms in terms of accuracy and detection speed. Although the algorithm based on deep learning takes up a lot of time in the early training and learning process, it takes less time in the later image detection, and the detection accuracy has been significantly improved. Fast r-cnn and the target detection algorithm studied in this article have very close accuracy, but the detection speed is about 7 times different. The fast r-cnn algorithm is divided into two parts, which consumes a lot of memory and time in storage and detection. This method combines the area designation operation with the convolutional network, which simplifies the network structure and improves the running speed. The deep learning target detection algorithm based on region nomination has high accuracy but slow speed. Although the speed can be increased by reducing the number of salient area suggestions or reducing the resolution of the input image, the speed has not been improved qualitatively.

The SSD target detection algorithm does not require area designation and is a direct regression location detection

algorithm. Due to the simple model, the algorithm is fast. Because there is no area designation operation, it is easy to miss the detection, so the accuracy is not as good as the deep learning target detection algorithm based on area designation. SSD target detection network has the problem of imbalance between positive and negative samples, which leads to poor detection of small targets and serious missed detection problems. The algorithm studied in this article improves the accuracy of target detection by constructing a multiscale feature network. By improving the VGG-Net model, the detection accuracy is improved and the training complexity of the model is reduced. Through the focus loss function, the ratio of positive and negative samples in the network is balanced, so that the detection performance of small targets in remote-sensing images is improved, and the overall missed detection rate of the algorithm is reduced. On the basis of ensuring the running time, the overall accuracy of the algorithm has been improved. In summary of the above two aspects of performance parameter research and analysis, the detection algorithm studied in this chapter has significantly improved the detection performance and has a clear advantage in the detection accuracy performance. At the same time, the neural network algorithm is used to speed up the processing speed through hardware equipment, and the real-time performance of target detection is improved.

5. Conclusions

Due to the limitations of the underwater environment, underwater navigation and positioning, target recognition, and underwater communication are usually realized by underwater acoustic information. Underwater sonar technology is one of the important technologies to be studied in the future ocean engineering. In this context, this article

conducts an in-depth study on the target segmentation method of underwater sonar images based on deep learning. Aiming at the sonar image's own characteristics such as large gray-scale distribution range, large amount of information, fuzzy boundary, and complex structure, this article mainly conducts an in-depth discussion from automatic threshold algorithm, fractal dimension algorithm, and algorithm based on Markov random field. This study discusses the advantages of the algorithm based on Markov random field and improves its segmentation algorithm. It proposes an automatic segmentation algorithm based on Markov and uses the measured sonar image to implement the algorithm. The main work and conclusion of this article are the pre-processing methods of sonar images, including denoising and gray-scale normalization methods. In the part of sonar image denoising, several commonly used denoising methods are discussed. By analyzing the characteristics of the single-side scanning sonar measurement image, the wavelet method is used to denoise, the algorithm principle, the closing value selection method, and the denoising steps are given, and the sonar image gray-scale normalization algorithm is given. The noise interference of the image is reduced, so that the gray-scale change range of the sonar image is kept consistent, which is convenient for subsequent image segmentation. This article discusses several commonly used sonar image segmentation methods in detail and in depth. Among several common target detection methods, for several traditional target detection algorithms, summarize their shortcomings in practical applications. The segmentation results of various algorithms are given, and their advantages, disadvantages, and existing problems are analyzed.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

- [1] R. Lou, W. Wang, X. Li, Y. Zheng, and Z. Lv, "Prediction of ocean wave height suitable for ship autopilot," *IEEE Transactions on Intelligent Transportation Systems*, 2021.
- [2] N. Krishnaraj, M. Elhoseny, M. Thenmozhi, M. M. Selim, and K. Shankar, "Deep learning model for real-time image compression in Internet of Underwater Things (IoUT)," *Journal of Real-Time Image Processing*, vol. 17, , 2020 In Press.
- [3] R. Lou, Z. Lv, S. Dang, T. Su, and X. Li, "Application of machine learning in ocean data," *Multimedia Systems*, pp. 1–10, 2021.
- [4] H. Zhu, H. Wei, B. Li, X. Yuan, and N. Kehtarnavaz, "Real-time moving object detection in high-resolution video sensing," *Sensors*, vol. 20, no. 12, 2020.
- [5] H. Wu, "Sonar image target detection and recognition based on convolution neural network," *Mobile Information Systems*, vol. 2021, no. 6, pp. 1–8, Article ID 5589154, 2021.
- [6] P. Tueller, R. Kastner, and R. Diamant, "Target detection using features for sonar images," *IET Radar, Sonar & Navigation*, vol. 14, no. 12, pp. 1940–1949, 2020.
- [7] M. Z. Lubis, C. Hu, H. Xu, and J. Cui, "Moving target detection for multi-beam forward-looking sonar based on frame-difference method," *Yi Qi Yi Biao Xue Bao/Chinese Journal of Scientific Instrument*, vol. 39, no. 2, pp. 169–176, 2018.
- [8] Q. Wang, Y. Li, and X. Liu, "Analysis of feature fatigue EEG signals based on wavelet entropy," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 32, no. 8, Article ID 1854023, 2018.
- [9] H. Chen, Y. Lu, and L. Tu, "Fault identification of gearbox degradation with optimized wavelet neural network," *Shock and Vibration*, vol. 20, no. 2, pp. 247–262, 2013.
- [10] X. Hao, G. Zhang, and S. Ma, "Deep learning," *International Journal of Semantic Computing*, vol. 10, no. 03, pp. 417–439, 2016.
- [11] L. June-Goo, J. Sanghoon, C. Young-Won et al., "Deep learning in medical imaging: general overview," *Korean Journal of Radiology*, vol. 18, no. 4, pp. 570–584, 2017.
- [12] K. Sirinukunwattana, E. Shan, Y. W. Tsang, R. J. Sned, I. A. Cree, and N. M. Rajpoot, "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1196–1206, 2016.
- [13] T. O Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Transactions on Cognitive Communications & Networking*, vol. 3, no. 4, pp. 563–575, 2017.
- [14] X. Wang, L. Gao, S. Mao, and S. Pandey, "CSI-based fingerprinting for indoor localization: a deep learning approach," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 1, pp. 763–776, 2016.
- [15] D. Ravi, C. Wong, F. Deligianni et al., "Deep learning for health informatics," *IEEE Journal of Biomedical & Health Informatics*, vol. 21, no. 1, pp. 4–21, 2017.
- [16] P. Burlina, K. D. Pacheco, N. Joshi, D. E. Freund, and N. M. Bressler, "Comparing humans and deep learning performance for grading AMD: a study in using universal deep features and transfer learning for automated AMD analysis," *Computers in Biology and Medicine*, vol. 82, pp. 80–86, 2017.
- [17] Y. Ji, Z. Jie, Y. Wang, and X. Chu, "Vessel target detection based on fusion range-Doppler image for dual-frequency high-frequency surface wave radar," *IET Radar, Sonar & Navigation*, vol. 10, no. 2, pp. 333–340, 2016.
- [18] J. Wang, G. Xin, F. X. Ge, and G. R. Arce, "Spatial coding in sonar imaging," *Electronics Letters*, vol. 53, no. 17, pp. 1222–1224, 2017.
- [19] R. T. Schirrmeyer, L. Gemein, K. Eggenberger, and F. Hutter, "Deep learning with convolutional neural networks for decoding and visualization of EEG pathology," *Human Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.
- [20] S. Ding, S. Qu, Y. Xi, and S. Wan, "Stimulus-driven and concept-driven analysis for image caption generation," *Neurocomputing*, vol. 398, 2019.
- [21] S. Wan, L. Qi, X. Xu, C. Tong, and Z. Gu, "Deep learning models for real-time human activity recognition with smartphones," *Mobile Networks and Applications*, vol. 25, pp. 1–13, 2019.
- [22] S. Albarqouni, C. Baur, F. Achilles, V. Belagiannis, S. Demirci, and N. Navab, "AggNet: deep learning from crowds for mitosis detection in breast cancer histology images," *IEEE*

- Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1313–1321, 2016.
- [23] W. Zhao and S. Du, “Spectral–spatial feature extraction for hyperspectral image classification: a dimension reduction and deep learning approach,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 8, pp. 4544–4554, 2016.
- [24] X. X. Zhu, D. Tuia, L. Mou et al., “Deep learning in remote sensing: a comprehensive review and list of resources,” *IEEE Geoscience & Remote Sensing Magazine*, vol. 5, no. 4, pp. 8–36, 2018.
- [25] L. D. Oliveira, M. Kagan, L. Mackey, and B. Nachman, “Jet-images — deep learning edition,” *Journal of High Energy Physics*, vol. 2016, no. 7, 32 pages, 2016.
- [26] C. A. Ronao and S. B. Cho, “Human activity recognition with smartphone sensors using deep learning neural networks - ScienceDirect,” *Expert Systems with Applications*, vol. 59, pp. 235–244, 2016.
- [27] M. Kallenberg, K. Petersen, M. Nielsen et al., “Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1322–1331, 2016.
- [28] G. B. Goh, N. O. Hodas, and A. Vishnu, “Deep learning for computational chemistry,” *Journal of Computational Chemistry*, vol. 38, no. 16, pp. 1291–1307, 2017.
- [29] Y. Chen, Z. Lin, Z. Xing, W. Gang, and G. Yanfeng, “Deep learning-based classification of hyperspectral data,” *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2094–2107, 2017.
- [30] S. Levine, P. Pastor, A. Krizhevsky, and D. Quillen, “Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection,” *The International Journal of Robotics Research*, vol. 37, no. 4-5, pp. 421–436, 2016.