WILEY | Hindawi

Research Article

Image Recognition Technology in Texture Identification of Marine Sediment Sonar Image

Chao Sun,[1,2] Li Wang ,[1] Nan Wang,[2] and Shaohua Jin[2]

[1]*Liaoning Normal University, Dalian, Liaoning, China*
[2]*PLA Dalian Naval Academy, Dalian, Liaoning, China*

Correspondence should be addressed to Li Wang; wang_li1125@126.com

Through the recognition of ocean sediment sonar images, the texture in the image can be classified, which provides an important basis for the classification of ocean sediment. Aiming at the problems of low efficiency, waste of human resources, and low accuracy in the traditional manual side-scan sonar image discrimination, this paper studies the application of image recognition technology in sonar image substrate texture discrimination, which is popular in many fields. At the same time, considering the scale complexity, diversity, multisources, and small sample characteristics of the marine sediment sonar image texture, the transfer learning is introduced into the image recognition, and the *K*-means clustering algorithm is used to reset the prior frame parameters to improve the speed and accuracy of image recognition. Through the experimental comparison between the original model and the new model based on transfer learning, the AP (average precision) value of the yolov3 model based on transfer learning can reach 84.39%, which is 0.97% higher than that of the original model, with considerable accuracy and room for improvement; it takes less than 0.2 seconds. This shows the applicability and development of image recognition technology in texture discrimination of bottom sonar images.

1. Introduction

Texture feature is the image feature with the most hidden information. Texture information shows different brightness or color in image features. In nature, no matter small objects or large-scale objects have certain texture distribution, which is a special internal correlation feature of objects [1]. Texture usually shows different gray distribution rules in sonar images, and this kind of distribution will show different information according to different situations, especially in the expression of marine sediment [2]. Because the marine sediment is often the same kind of material in a large range, the texture of the sediment in the area may represent a kind of microtopography in the sonogram image. It may also be a form of existence that cannot be displayed in the terrain but can have an impact on human activities; if this kind of texture features can be detected and recognized, the sediment information can be classified according to this kind of texture features. Furthermore, it can express more information about

the substrate in addition to its composition and structure. The side-scan sonar image can be used to distinguish the substrate texture, but the traditional sonar image discrimination method is too dependent on subjective consciousness, so the accuracy is not very high [3]. This paper presents the application of image recognition technology in texture discrimination of bottom sonar image. This study can greatly improve the weaknesses of low efficiency, low accuracy, and strong subjective consciousness of artificial discrimination of marine sediment side-scan sonar so as to provide sufficient data support and decision-making basis for further classification of marine sediment.

2. Background and Related Work

2.1. Convolutional Neural Networks. The convolutional neural network, CNN, is a class that contains convolution computation and has depth structure of feedforward neural

networks (FNNs) [4, 5]. It is one of the representative algorithms of deep learning [6].

The convolution neural network structure consists of input layer, hidden layer (convolution layer, pooling layer, and full connection layer), and output layer [7]. As shown in Figure 1, it is a schematic diagram for the operation of the convolution neural network.

The main task of target detection technology is to locate the target of interest from the image, which needs to accurately judge the specific category of each target and give the boundary box of each target, which is more complex and interesting than the simple convolution neural network image classification. In recent years, the target detection technology based on the convolution neural network has been significantly improved [8]. The more popular algorithms can be roughly divided into two categories, one is the R-CNN (region convolutional neural network) algorithm based on region proposal (R-CNN, Fast R-CNN, Faster R-CNN), in which two stages first get the candidate regions, then classify the candidate regions, and then carry on the frame regression. The other is the one stage algorithm such as YOLO (you only look once) and SSD (single-shot multibox detector), which uses only one CNN network to directly predict the categories and locations of different targets [9]. The first kind of method is more accurate, but the speed is slow; the second kind of algorithm is faster, but the accuracy is lower. The R-CNN series model automatically recognizes the target of side-scan sonar image with long time and low efficiency, and it belongs to small sample database for side-scan sonar image. The single detection network such as YOLO generalizes the target detection problem into a regression problem and has a simple network structure, and the detection speed is faster than the former, which can basically meet the requirements of real-time detection. Therefore, this paper proposes an improved yolov3 model side-scan sonar substrate image texture feature recognition based on transfer learning.

2.2. YOLO Model. The YOLO model is an object recognition and location algorithm based on the depth neural network. Its biggest feature is its fast-running speed and can be used in real-time systems [10]. Different from other convolution neural network classifiers using sliding windows, YOLO integrates target location, target region frame prediction, feature extraction, and classification into a single neural network model to realize end-to-end target recognition based on the deep convolution neural network [11]. The whole training and detection process, data input, and result output of YOLO are completed in the network, so it has better accuracy and faster recognition speed. The YOLO network consists of 24 convolution layers and 2 fully connected layers, in which the convolution layer is used to extract image features, and the fully connected layer is used to output prediction probability and location coordinates. In addition, YOLO uses half of the resolution to preprocess the convolution layer for training and doubles the resolution for target detection, so as to achieve the purpose of fast detection [12]. The recognition process of the YOLO algorithm is roughly divided into three steps: first, unify the input image size to 448-448-102, then run the CNN algorithm to train or test the image, and finally optimize the detection results by nonmaximum suppression (nonmaximum suppression, NMS), such as errors! The reference source was not found. As shown in Figure 2, the main improvement of the yolov3 network compared with the YOLO model is that it adjusts the network structure, proposes a new backbone network Darknet-53, and constructs the multiscale feature of FPN (feature pyramid network) support for detection.

The specific model structure is shown in Figure 3. The yolov3 model divides the input images into $S \times S$ grids and extracts full convolution features through the Darknet-53 basic network deepened by the residual network [13]. At the same time, the up-sampling and fusion methods similar to FPN are used to detect on multiscale feature maps, which makes the new network improve the detection accuracy while keeping the speed advantage, especially the detection ability of multiscale targets.

3. Improvement of K-Means Clustering for Prior Frame

Based on the current data and data support, this paper intends to cluster and correct the prior frame again and then introduces transfer learning to increase the learning data according to the small amount of sonar image data, so as to improve the accuracy and calculation speed and maximize the image recognition advantages of the yolov3 model [14].

According to the network detection mechanism, the selection of prior frame parameters has a direct impact on the recognition accuracy. The nine preset anchor frames in the original yolov3 are obtained by clustering COCO data, while the COCO data set contains 80 categories, and the scale, shape, and size are relatively evenly distributed [15]. However, the size of the substrate texture target range of the side-scan sonar image in this experimental data set exists.

The prior bounding box is preferably available in various sizes and can focus on small size, so it is necessary to improve the prior box of the original COCO data set. In this paper, the K-means clustering algorithm is used to recluster the sunken ship data set. After six times clustering, the clustering results tend to be stable.

In this data set, there are six sizes (13 22), (31 51), (52 88), (75 95), (95 110), and (113 162). As shown in Figure 4, the original prior frame cannot well adapt to the substrate texture target of side-scan sonar image, while the reclustered prior frame is more consistent with the shape characteristics of sediment texture target.

4. Cite Transfer Learning in the Experiment

With the emergence of more and more machine learning application scenarios and the existing better supervised learning requires a lot of tagging data, tagging data is a boring and costly task, so transfer learning has received more and more attention [16]. Transfer learning applies the knowledge or patterns learned in a certain field or task to different but related fields or problems, transfers labeled data
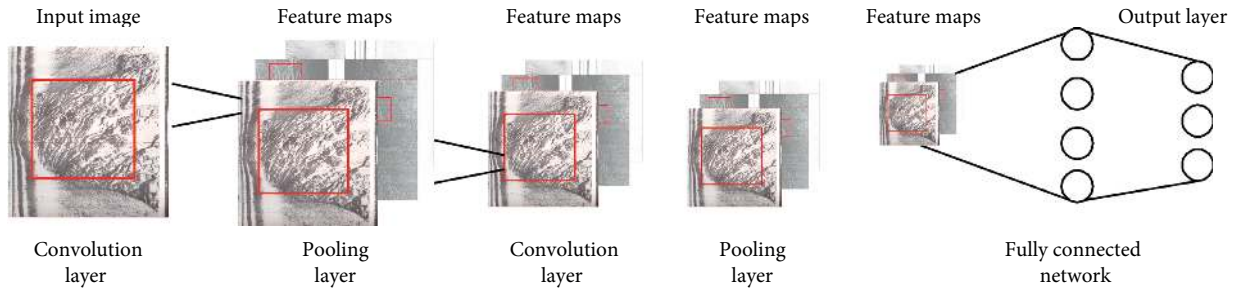
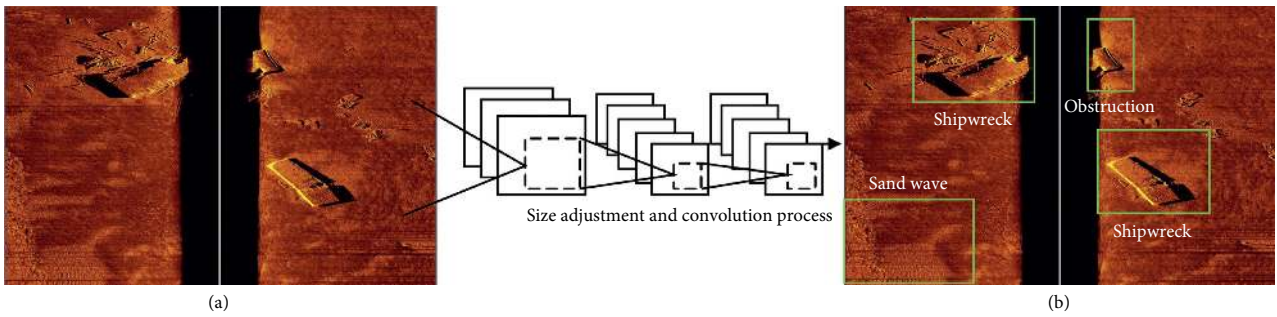Figure 1: Schematic diagram of convolution neural network operation.

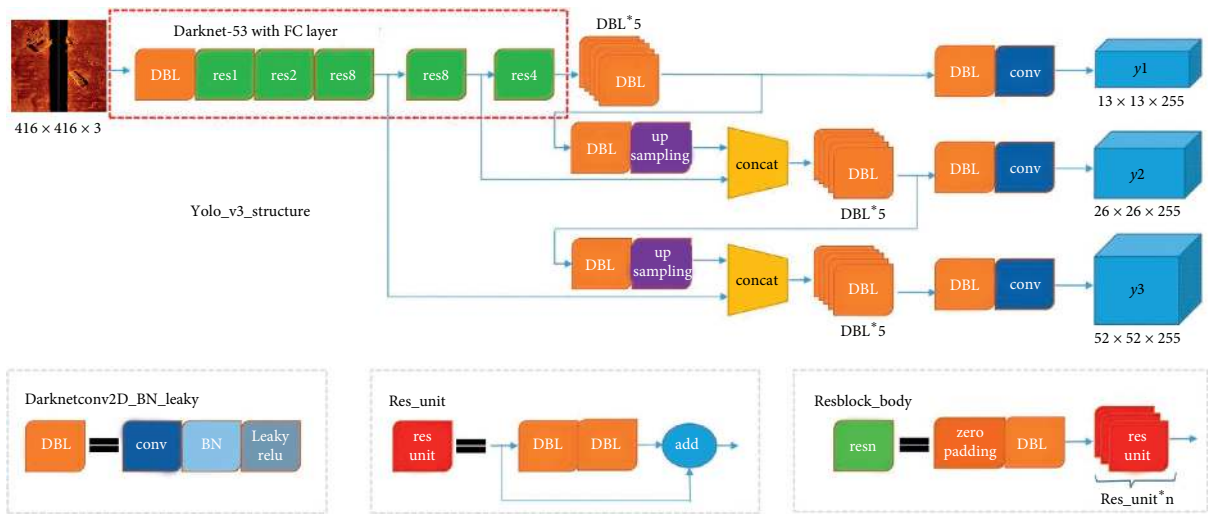Figure 2: YOLO model flow of marine sediment.

Figure 3: Example diagram of the range of action of a priori box on a substrate texture target.
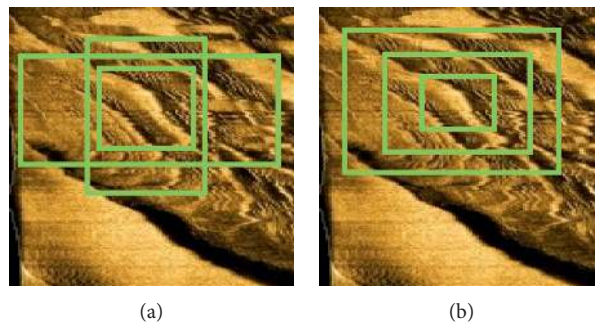

Figure 4: An example of the range of action of a priori box on a substrate texture target.

or knowledge structure from related fields, and completes or improves the learning effect of the target domain or task.

The knowledge gained from previous tasks can be directly applied to new tasks through prior transformation or even minimal transformation. When these effective features are captured and applied to new tasks, they are applied to transfer learning [17]. In view of the small number of texture feature samples of side-scan sonar, transfer learning can share the learned model parameters with the new model through migration to speed up and optimize the learning efficiency of the model, reduce repetitive work, rely on target task training data, and improve the performance of the model [18]. The specific flow chart is shown in Figure 5.

5. A Comparative Experiment between the Original Model and the Transfer Learning Model

The environment for experimental training and testing in this paper is shown in Table 1.

The experimental data of this paper are composed of 1000 side-scan sonar images and pictures provided by side-scan sonar manufacturers and related types of soil pictures. In this paper, the waterfall map of side-scan sonar is processed, and then all the pictures are standardized and modeled. The pixel of the picture is fixed as 416 (pixel) × 416 (pixel), and the preprocessed data set has a total of 3000.

After statistics and analysis of the data set, 2000 are taken as the training set and the remaining 1000 are used as the test data set of the experiment. Among them, the training is concentrated, with a total of 1000 steps. The loss values of the original yolov3 model and the improved yolov3 model based on transfer learning are shown in Figure 6. The loss values of the two models decrease with the increase in training times and tend to be stable after a certain number of times, which also shows that the target detection model is very applicable and feasible in the texture feature recognition of marine sediment sonogram. After learning the original yolov3 model, the region is stable after 400 steps of training, and the loss value is finally stable at about 7.8. In the improved yolov3 model of transfer learning, because some of the shallow feature extraction parameters of the model are completed based on the COCO data set, the initial loss value decreases rapidly; although some features of the side-scan sonar texture feature data set and COCO data can be integrated, there are many differences and dominate learning, so there are still great fluctuations after the loss value decreases quickly. And there are still two big fluctuations after 400 steps. However, there is a model that can well obtain the multiscale information and location information of the target, so the improved yolov3 model based on transfer learning is gradually stable in the region after two large fluctuations, and the final loss value tends to converge after 530 steps of training, which is about 6.7, which is lower than that of the original Euclidean yolov3 model. It is proved that

the improved yolov3 model based on transfer learning has better learning and adaptability to new objects [19].

The evaluation criteria used in this experiment are average accuracy (AP average precision) and harmonic average $F1$. The term AP actually comes from the field of information retrieval [20], which reflects the performance of the whole model. It is the area value of P-R (precision-recall) curve, that is, the average accuracy [21]. Among them, precision (accuracy rate of preschool education) indicates how many of the detected targets are accurate and an index to measure the accuracy of the results. Recall (recall rate, also known as recall rate) indicates how many accurate targets have been detected, which is used to measure the integrity of the results [22]. According to the classification results, the classified samples can be divided into four categories: correctly classified positive samples (TP true positives), misclassified positive samples (FP false positives), correctly classified negative samples (TN true negatives), and misclassified negative samples (FN false negatives). TP+FP is the total number of classified samples, and TP+FN is the total number of positive samples. Formulas (1) and (2), for example, represent the formulas of accuracy and recall:

$$P\,(\text{precision}) = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{1}$$

$$R\,(\text{recall}) = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{2}$$

Therefore, according to the definition of the four types of samples, the average accuracy, that is, the evaluation formula of AP is shown in the following formula:

$$\text{AP} = \int_0^1 P\,(R)\text{d}R. \tag{3}$$

According to the AP formula, the original yolov3 model and the improved yolov3 model based on transfer learning are used to detect the same set of data sets in the test environment, and the two models are obtained. The larger the area of the curve and axis is, the higher the AP value is and the better the detection effect of the model is.

The P-R diagrams of the original yolov3 model and the improved yolov3 model based on transfer learning are shown in Figures 7(a) and 7(b), respectively. As can be seen above, the area of the curve and axis represents the worth of the AP [23]. Comparing the P-R curve and AP value of the two models, the AP values are 83.42% and 84.39, respectively. It can be seen that the AP value of the improved yolov3 model based on transfer learning is significantly higher than that of the original yolov3 model without improvement, and the average accuracy is higher than 0.97. It is not difficult to see from the figure that the accuracy of the original structure yolov3 model is close to 80% when the recall rate reaches 82%, but with the further improvement of the recall rate, the accuracy is declining. This shows that when the data set increases and the recall increases, the
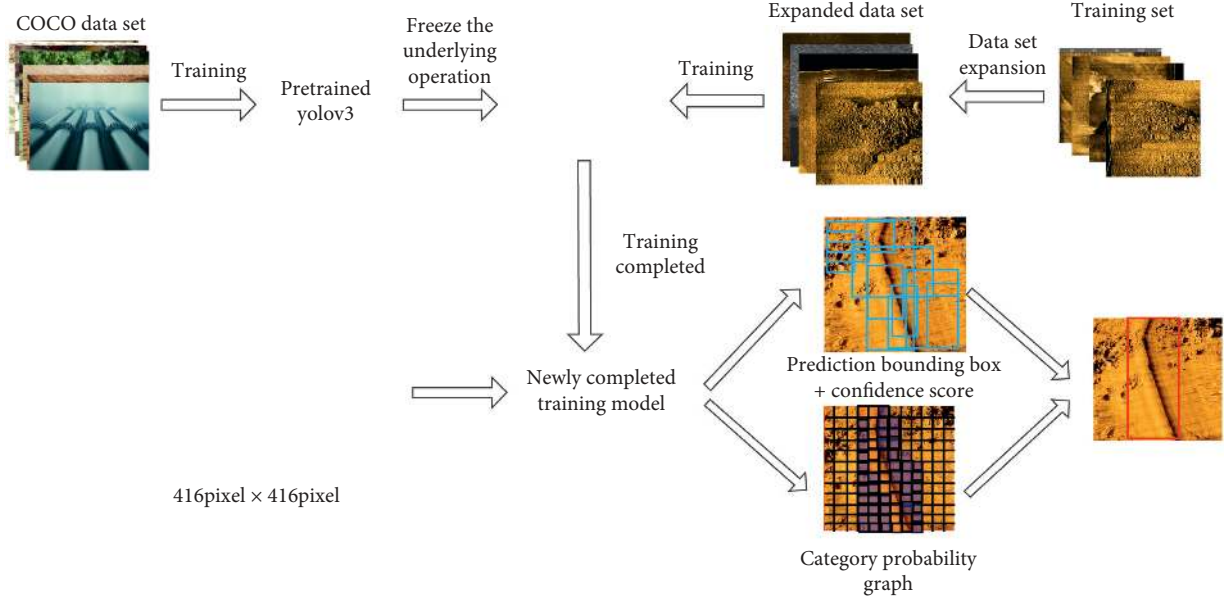
Figure 5: Flow chart of transfer learning.

accuracy of the original structure yolov3 model cannot provide the function of high accuracy. However, the improved yolov3 model based on transfer learning has almost no decline in the curve trend while maintaining a higher accuracy [24]. With the improvement of recall, the accuracy still has an upward trend, gradually approaching the accuracy of more than 90%. This shows that the yolov3 model with transfer learning can maintain high accuracy in the case of high recall, and the improved yolov3 model can overcome the disadvantage of rapid decline in the accuracy of the original yolov3 model. This also proves that the improved yolov3 model based on transfer learning is very effective in the recognition of side-scan sonar texture targets, and the model can maintain close to 85% recognition effect when the recall rate is more than 85%.

Accuracy and recall are actually a pair of contradictory measures. Generally speaking, when the accuracy is high, the recall rate is often on the low side, while the recall rate is high, the accuracy rate will have a downward trend [25]. However, for the evaluation of an algorithm, it is impossible to consider only one aspect of performance, so it is necessary to comprehensively consider the performance measurement of accuracy and recall, so the $F1$ metric is introduced [26]. $F1$ measure is the harmonic average of accuracy and recall. The $F1$ value obtained by formula (4) is used to represent the comprehensive performance of the algorithm. In this paper, the interval confidence and IOU are set to 0.5. The test results of the two models are shown in Table 2.

$$F1 = 2 \times \frac{2 \times P \times R}{P + R}. \tag{4}$$

As can be seen from Table 2, the recognition accuracy of the original yolov3 model can also reach about 85%, but the accuracy of the improved yolov3 model based on transfer

Table 1: Experimental environment.

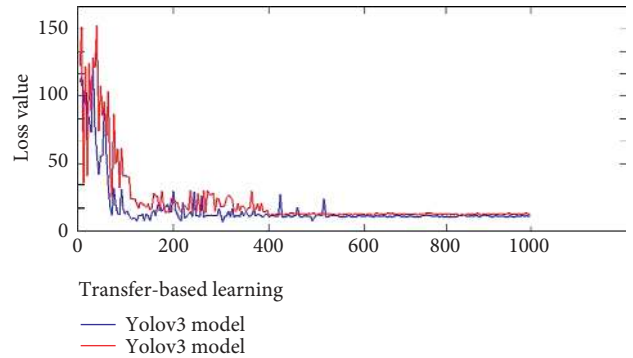| Name	Configuration
Memory	8 GB
CPU	Inter(R) Core(TM) i7-7700 CPU@ 3.60 ghz
Deep learning framework	Tensorflow
Operating system	Windows 10
GPU	NVIDIA TITAN RTX

Figure 6: Two yolov3 models loss value.

learning is 3.24% higher, and the latter can also be 2.28% higher in recall. It can be seen that the improved model is suitable for the recognition of substrate texture. At the same time, the values of AP and $F1$ are improved in varying degrees after transfer learning, but the detection time is not increased, which shows that although there are many and large data sets in the early learning process when transfer learning is introduced, this does not affect the target

Class: 83.42% = target AP
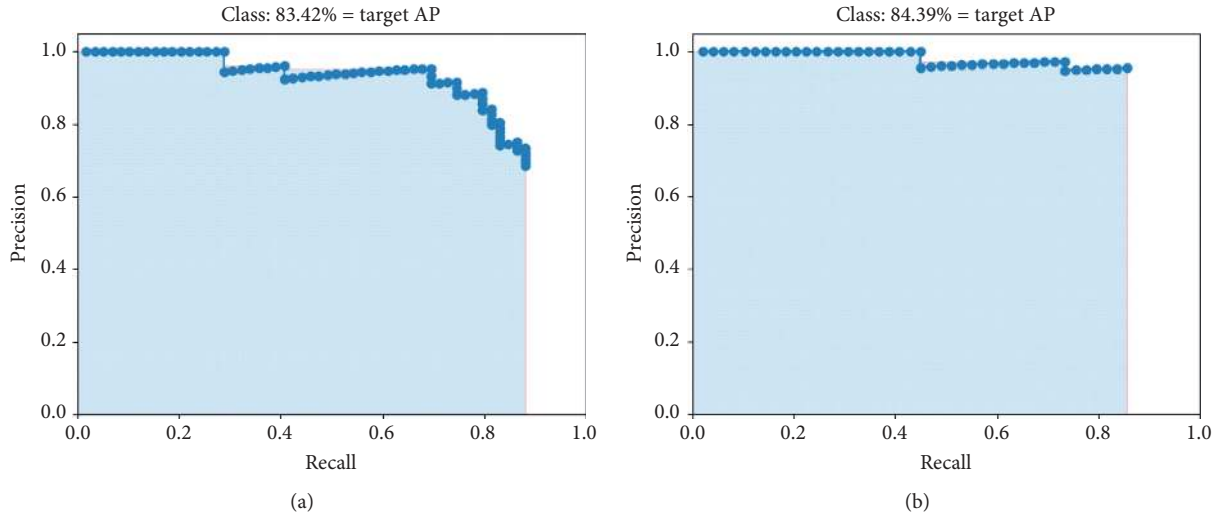
Class: 84.39% = target AP

(a)

(b)

Figure 7: The P-R curves of the two models: (a) original model and (b) yolov3 model based on transfer learning.

Table 2: Comparison of test results of two models.

Model	Precision (%)	Recall (%)	AP (%)	F1 (%)	Time (s)
yolov3 model	85.54	83.38	83.42	84.45	0.17
yolov3 of transfer learning	88.78	85.66	84.39	87.19	0.17

recognition work in the later stage, and the new model can also achieve an excellent performance of 0.17 s in the case of overall improvement in detection accuracy and recall. The detection time is not affected, and it can maintain high recognition efficiency and get more accurate results in the same time. It is proved that the comprehensive performance of the improved yolov3 model based on transfer learning is due to the original model, and all aspects of the index are improved in varying degrees, which can adapt to the recognition of targets such as marine sediment texture features [27].

In the process of detection, the detection effect of the two models in marine geological texture feature target recognition can also be analyzed through the detection results of the detection image. Among the three groups of graphs, the (a) diagram of each group represents the detection results of the original structure YOLOv3 model, and the (b) diagram of each group represents the detection results of the improved YOLOv3 model based on transfer learning. As shown in Figure 8, the characteristic of this image is that there are not only a wide range of marine sediment texture features but also small area texture features, which tests the ability of the model to detect texture feature targets with different scales and regions. It can be seen that in the an image, the original structure yolov3 model does not detect a small part of the texture features of the ocean bottom, while the red box in the *b* map is the missing part of the original

model [28]. The existence of targets of different scales is an important test for the image recognition model. If too many targets are missed, the reference significance of the recognition results will be affected. Although the area in the image is small, it may be a large area in the actual seafloor, which will have a certain confusing impact on human production activities, military operations, and so on [29].

Figure 9 is a comparison of the detection results for the recognition of marine sediment texture feature images in a large area. It can be seen that in the recognition results of (a) and (b), the large-scale texture feature targets on the right side are accurately recognized, which shows that in large-scale image recognition, the two models have high accuracy, and the improved yolov3 model is not affected by modification or inherits the advantages of large-scale target recognition in the original model.

Figure 10 shows a submarine uplift texture feature. It is not difficult to see from Figure 10(a) that although the original structure yolov3 model can accurately find most of the structure of the target, the positioning accuracy is not high. The marine sediment uplift of the nearby area is also included in the target, so that the location area is much larger than the actual area, but the improved yolov3 model based on transfer learning can identify and locate the target with higher accuracy. Other typical texture features are not put into the recognition box to confuse the target.
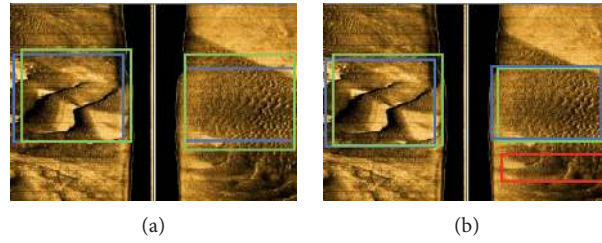
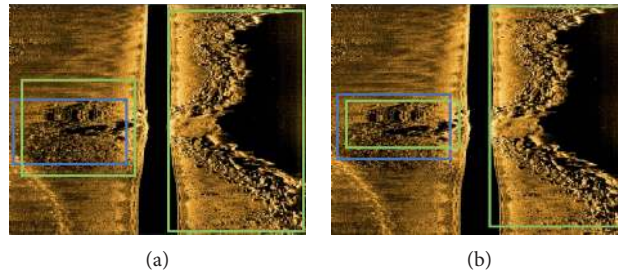Figure 8: Comparison of target detection results at different scales.

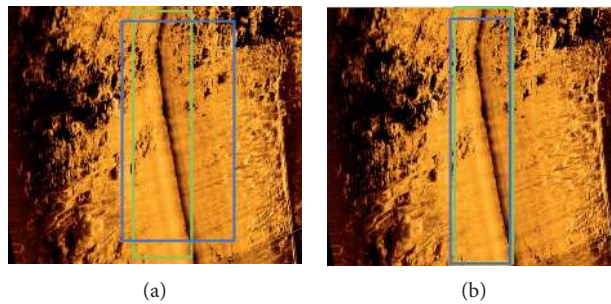Figure 9: Comparison of detection results of texture features in large area.

Figure 10: Comparison of positioning accuracy in test results.

6. Conclusion, Limitations, and Future Work

Through the comparison of three groups of detection images, it can be seen that the yolov3 model based on transfer learning shows better comprehensive performance than the original model in marine sediment texture feature recognition. No matter the average accuracy or $F1$ value or the comparison in the detection results, we can see its excellent comprehensive performance in marine sediment texture feature recognition, such as low leakage alarm rate, high accuracy, and so on. Both the original yolov3 model and the yolov3 model with transfer learning have high accuracy and high recognition efficiency, which shows that the image recognition technology has great efficiency and accuracy advantages in the texture discrimination of ocean sediment sonar images, and the effect is better after the transfer learning and the improvement of the prior frame. If more samples can be introduced into the future research and more improvements can be made at the same time, the author believes that the image recognition technology has more advantages over the sonar image recognition.

Image recognition technology has been widely used in various disciplines and work, but it is rarely used in the side-scan sonar image recognition of marine sediment, and the original image recognition technology is more often used. The author introduces the improved yolov3 model based on transfer learning into it although it can improve the efficiency and accuracy of side-scan sonar substrate sonogram recognition, but the success rate of image recognition is still the focus of the later work, and the error rate should be reduced to a minimum so as to provide more powerful support for further detection and classification work, so the recognition accuracy is still the focus of the future work. The authors hope that it can be solved in the future research.

Data Availability

The data generated or analyzed during this study are included within this article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] X. Liu, "Reconstruction of the south China sea coral reef shipwreck and surrounding topography based on side scan sonar images," *Tropical Geography V*, vol. 40, no. 02, pp. 106–116, 2020.

[2] M. Rahnemoonfar, A. F. Rahman, R. J. Kline, and A. Greene, "Automatic seagrass disturbance pattern identification on sonar images," *IEEE Journal of Oceanic Engineering*, vol. 44, no. 1, pp. 132–141, 2019.

[3] W. Zeng, "Dynamic model simulation of long-term evolution of seabed terrain of dumping area based on seabed side-scan sonar detection technology," *Journal of Coastal Research*, vol. 83, pp. 23–28, 2018.

[4] J. Gu, "Recent advances in convolutional neural networks," *Pattern Recognition*, vol. 77, pp. 354–377, 2015.

[5] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*, MIT press, Cambridge, MA, USA, 2016.

[6] W. Zhang, "Shift-invariant pattern recognition neural network and its optical architecture," in *Proceedings of Annual Conference of the Japan Society of Applied Physics*, Hamamatsu, Japan, June 1988.

[7] A. Ng, K. Kian, and B. Younes, "Convolutional neural networks, deep learning. Coursera and deeplearning.AI," 2018.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proceedings of the International Conference on N euralInformation Processing System*MIT Press, Cambridge, MA, USA, June 2012.

[9] M. Lin, Q. Chen, and S. Yan, "Network computer science," 2013.

[10] R. Webster, *Geostatistics for Environmental Scientists*, Wiley, Hoboken, NY, USA, 2nd edition, 2002.

[11] E. J. Wood, "Practical statistics for environmental and biological scientists," *Biochemistry and Molecular Biology Education*, vol. 30, no. 4, pp. 272-273, 2002.

[12] C. Szegedy, W. Liu, and Y. Jia, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, June 2015.

[13] W. Liu, "SSD: single shot multibox detector," in *Proceedings of the European Conference on Computer Vision*, Springer International Publishing, Amsterdam, The Netherlands, October 2016.

[14] T. M. Mitchell, *Machine Learning*, McGraw-Hill, New York, NY, USA, 2003.

[15] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

[16] D. Haritha, "An inductive transfer learning approach using cycle-consistent adversarial domain adaptation with application to brain tumor segmentation," 2020, https://arxiv.org/abs/2005.04906.

[17] R. Girshick, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Columbus, OH, USA, June 2014.

[18] H. Zhichao and G. Zhang, "Dog recognition algorithm in elevator car based on YOLO," *Industrial Control computer*, vol. 32, no. 6, pp. 15-16, 2019.

[19] Y. I. Shi, "Nighttime target recognition method based on infrared thermal imaging and YOLOv3," *Infrared Technology*, vol. 41, no. 10, pp. 970–975, 2019.

[20] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes challenge: a retrospective," *International Journal of Computer Vision*, vol. 111, no. 1, pp. 98–136, 2015.

[21] X. Wei, D. Wei, D. Suo, L. Jia, and Y. Li, "Multi-target defect identification for railway track line based on image processing and improved YOLOv3 model," *IEEE Access*, vol. 8, pp. 61973–61988, 2020.

[22] T. Y. Lin, "Feature Pyramid networks for object detection," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE Computer Society, Honolulu, HI, USA, July 2017.

[23] V. S. Cherkassky and F. Mulier, *Pattern Recognition*, IEEE Press, New York, NY, USA, 2017.

[24] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in *Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition*, pp. 6517–6525, IEEE, 2017.

[25] J. Redmon and F. Ali, *Computer Vision and Pattern Recognition*, pp. 101–110, IEEE Press, New York, NY, USA, 2017.

[26] M. J. Park and B. C. Ko, "Two-step real-time night-time fire detection in an urban environment using static ELASTIC-YOLOv3 and temporal fire-tube," *Sensors (Basel, Switzerland)*, vol. 20, p. 8, 2020.

[27] H. Wang, B. Zhou, and Y. Pan, "High resolution optical remote sensing image based on YOLOv3 network for classification and detection of ships on the sea," *Science and Technology Bulletin V*, vol. 36, no. 259, pp. 47–62, 2020.

[28] S. Stanic, R. R. Goodman, K. B. Briggs, N. P. Chotiros, and E. T. Kennedy, "Shallow-water bottom reverberation measurements," *IEEE Journal of Oceanic Engineering*, vol. 23, no. 3, pp. 203–210, 1998.

[29] J. Waldemark, "Hybrid neural networks for automatic target recognition," in *Proceedings of the IEEE International Conference on Systems*, IEEE, Orlando, FL, USA, October 1997.