Computer Fraud and Security
ISSN (online): 1873-7056

Research on Underwater Optical Image Target Detection and Network Security Protection Method based on Deep Learning

KuoChai, HeyuWang
Liaoning Technical University
Corresponding author email：wanghy_75@163.com

Abstract

This paper studies underwater optical image target detection and network security protection methods based on deep learning. The YOLOv7-based method is used to solve the color distortion and blur caused by the scattering and absorption of light. By replacing the ELAN structure in the backbone network with the NewDSCLayer structure, the ELAN structure in the detection head module with the GSCLayer structure, and introducing the SimAM attention mechanism, the detection accuracy of overlapping targets and small targets is effectively improved, and the detection speed is also improved. In addition, this paper also designs an adaptive network security protection strategy based on deep reinforcement learning, which can identify multiple network threats through behavior analysis and feature matching, and dynamically adjust defense measures. Experiments have shown that the algorithm proposed in this paper has higher accuracy, faster response speed and higher system stability, and can provide efficient protection for network security.

Keywords : deep learning; optical image; target detection; network security protection

Introduction

Due to the complexity of the underwater environment and the characteristics of light scattering and absorption, the images collected by underwater vehicles have problems such as color distortion and low resolution. This degradation of imaging quality directly affects subsequent target recognition, thereby causing a decrease in the operating performance of the AUV [1]. Therefore, how to effectively improve the quality of underwater imaging and improve color restoration and clarity is the key to improving the level of underwater environmental perception and the work efficiency of robots. Using image enhancement methods to process the original image can not only solve problems such as color distortion, blur and insufficient contrast, but also provide high-quality environmental perception data for subsequent target detection.

At the same time, network security issues are becoming more and more prominent. With the increasing number of network attacks such as malware, botnets, and phishing, their concealment and destructive power are also increasing. This not only endangers people's privacy and disrupts people's normal lives, but also has a great impact on the company's operations and even the country's security. In such an environment, it is particularly important to establish a flexible and adaptable information security protection system [2]. The defense system must be able to adapt to various new threats, and the system can promptly and effectively defend against various types of network attacks, ensuring the stability and security of the network environment.

1 YOLOv7 object detection algorithm

1.1 Input Module

The input module resizes the image according to the size of the backbone network.In addition, there are other ways to use the input layer, which will only increase the overhead of training samples, but will not affect the accuracy of detection. First, based on CutMix, the fused Mosaic data is enhanced and randomly arranged to form a larger fused image. Its advantage is that it can improve the model's adaptability to objects of different positions, sizes, directions, etc., increase data diversity, and improve the model's generalization ability. In addition, Mosaic data enhancement technology can also reduce GPU storage space usage and improve training efficiency.

1.2 Backbone module

The model includes several SConv convolutions, ELAN convolutions, and MPConv convolutions. Among them, the SConv convolution network is the bottom layer of YOLOv7 and other architectures[3]. In the convolution network, the convolution network has several different convolution kernels and strides. Usually, a convolution layer with a 1x1 convolution kernel is changed the number of communications; a convolution network with a 3x3 convolution kernel and a 1-level stride is usually used as a feature extraction algorithm; downsampling usually uses a 3x3 convolution kernel and a 2-order convolution neural network.

Figure 1 SConv layer structure

As shown in Figure 2, the conventional ReLU activation function is actually a maximum value function, with fast calculation speed, unsaturated gradient, and fast convergence speed, but some neurons cannot be fully activated, which affects the convergence speed of the network and is very sensitive to parameter initialization and learning rate. Therefore, the YOLOv7 network chooses a smoother non-monotonic SiLU as the activation function.

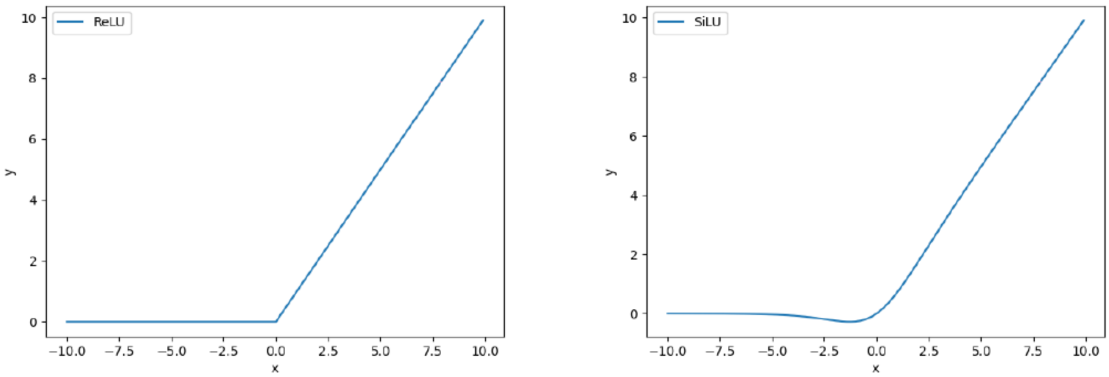

Figure 2 Comparison of SiLU activation function and ReLU activation function

The ELAN convolutional layer is a scalable and efficient hierarchical aggregation network. It improves the learning performance of the original network by transforming, expanding, and fusing cardinalities. Its greatest advantage is that it can avoid destroying the stability of the original gradient trajectory due to too many iterative units. The convolution structure of the ELAN component is a dual path. The former is relatively simple, and only

1x1 convolution is needed to adjust the number of feature paths. The second path is a more complex path, which first uses 1x1 convolution to transform the number of feature paths, and then performs feature extraction. This dual branch structure can efficiently extract feature information on the basis of realizing flexible adjustment of feature channels[4]; on this basis, it performs four 3×3 convolution operations for feature extraction. As can be seen from Figure 3, the E-ELAN structure finally superimposes these four characteristics to obtain the final feature extraction result. The E-ELAN convolutional layer expands the channels and cardinality of the operation module in a group convolution manner. While maintaining the transformation layer structure, it only changes the characteristics of the operation module itself, thereby effectively strengthening the feature learning based on multiple feature maps, improving the model's operational efficiency, and enhancing the robustness of the algorithm.

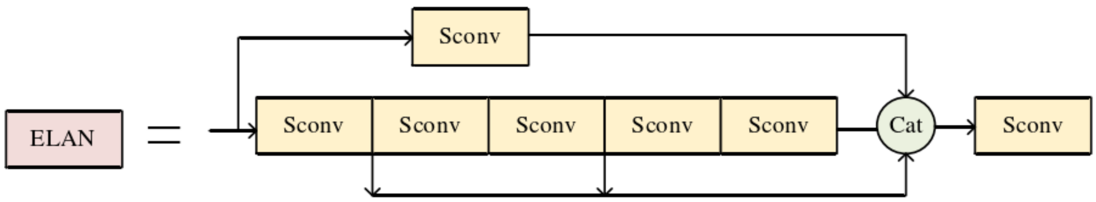

Figure 3 ELAN layer structure diagram

The MPConv convolution layer is a layer that adds a Maxpool layer to the SConv layer, forming two branches, the upper and lower branches. As can be seen from Figure 4, the upper branch passes through the Maxpool layer, which is the maximum pooling layer. Its function is to downsample the length and width of the image, reducing the length and width of the image by half, and passes through a layer of SConv layer with a convolution kernel of 1x1 to reduce the channels of the image by half [5]. The image is further segmented using a 3x3 convolution kernel and a 2nd-order SConv layer, and its length and width are reduced by half respectively; By using the "merge" operation, the feature information in the two paths is integrated together, thereby improving the expressive power of the model. The features of the upper and lower branches are connected in series to obtain a more discriminative feature expression, thereby improving the expressive power of the entire network.

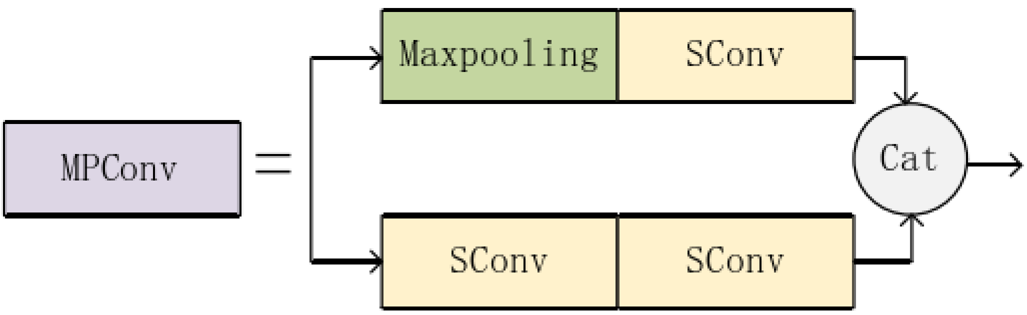

Figure 4. MPConv structure diagram

1.3 Head module

In terms of feature fusion, this paper organically integrates multi-scale elements to establish a feature pyramid with three levels of large scale, medium scale and small scale. The detection front end of YOLOv7 adopts the "path aggregation" feature cone (PAFPN) in the YOLO series, and realizes the detection of multi-scale objects through cross-level connectivity and feature fusion. The structure mainly consists of a SPPCPC layer, multiple SConv layers, an MPConv layer, and an upsampling layer. Through the bottom-up information transmission mechanism, low-level features are effectively transmitted to high-level layers, achieving deep fusion of multi-level features[6]. Among them, the SPPCPC layer enhances the network's perception of visual information through parallel multi-scale pooling operations (MaxPool), solves the limitations of convolutional neural networks in extracting repeated features, and thus significantly improves the generalization performance of the model.

1.4 Prediction Module

The prediction model adopts the REP structure and adjusts the number of image channels at three scales, and uses a 1x1 convolutional network to predict confidence, category, and anchor box. The REP structure consists of three branches. The top layer is a 3x3 convolutional layer for adjusting extraction; the middle branch is a 1x1 convolutional network for smoothing features; the lowest branch is the BN level, which has no special meaning.

2 Underwater optical image target detection based on deep learning

In view of the fact that underwater images are mostly randomly distributed and overlapped, this paper replaces the original ELAN model with the conventional convolutional layer in the ELAN model of the original network backbone network, and replaces the original ELAN structure with the NewDSCLayer structure. The deep separable convolutional layer with a large convolutional core is used to improve the receptive field, improve the detection accuracy of overlapping objects, and improve the detection speed.

2.1 Backbone Network Improvement

(1) Use deep separable large convolution kernels instead of standard convolution

Although the depthwise separable convolution (DSConv) method can effectively reduce the number of parameters in the model, there are still some errors. Experiments show that when DSConv has a 5x5 convolution kernel, the algorithm can well balance detection speed and detection accuracy[7]. To this end, using deep neural networks for convolution, the depth is increased from 3 × 3 to 5 × 5. This algorithm can not only reduce the size of the model and speed up the detection process, but also increase the effective detection area, thereby improving the detection accuracy.

(2) Deep Separable Convolution Transformed into Residual Network

Theoretically, for a general network, the higher the learning depth, the stronger its learning ability. But in fact, as the depth increases, the algorithm training difficulty will also increase, which means that the training error will become larger and larger, and the final accuracy will become lower and lower.

A skip connection is added between the input and output, and the output of the convolutional layer is added to the final output shown in Figure 5. Therefore, when the network is transmitted in reverse, the skip connection can directly transmit the gradient of the loss to the previous network, alleviating the problem of network performance degradation [8]. This residual structure allows the convolutional network to learn only a small part of the residual between the input and output instead of all input features, making it easy to train. Secondly, by using the identity mapping method, the problem of gradient loss is achieved when the gradient is transmitted in reverse, and in the convolution learning of a certain layer, when the weight is zero, it can still be passed to the previous layer. In this way, the separable deep convolution can be converted into a residual neural network, thereby ensuring the training process of the neural network.

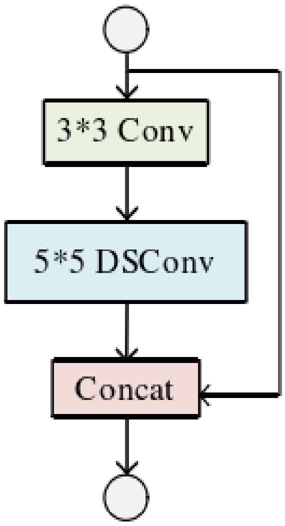

Figure 5 DSCBlock structure diagram

(3) Introducing an initialized feature extraction method at multiple scales

Figure 5 , since there is only one convolution layer in the network, the target features extracted in the network will be insufficient, this will lead to a decrease in detection accuracy.. Therefore, the structure of the residual neural network will be further improved to improve the accuracy of target detection. A new branch is added with a 1x1 convolution layer.

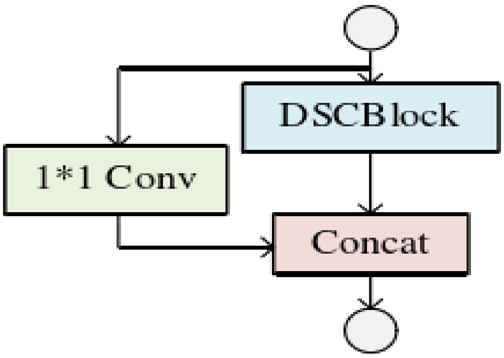

Figure 6 I-DSCBlock structure diagram

As shown in Figure 6, I-DSCBlock is an improvement on DSCBlock. It adds a 1*1 convolution on the left side of the structure, performs an addition operation on it, and then concatenates the feature maps of the two regions to obtain more feature information.

(4) Introducing channel attention to the end of the network to improve detection accuracy

Compared with the original ELAN structure, the new DSCLayer structure has a simpler structure, so it has a higher detection speed for objects in the entire structure; in addition, the introduction of large convolution kernels and the application of attention mechanism can also improve the accuracy of detection. Using the NewDSCLayer structure to replace the ELAN structure on the backbone network smaller improvements can be made to the object detection accuracy,without increasing the computational complexity of the model.

2.2 Improvement of detection head

To improve prediction efficiency, images in convolutional neural networks must undergo a series of similar transformations, that is, spatial information is transferred to channels step by step. When the feature map is expanded in space (length, width) and channels, some semantic information is often lost. The channel dense convolution algorithm maximizes the connection between channels, while the channel sparse convolution can completely cut off the connection between channels. The lightweight separable convolution algorithm (GSConv) can maintain network connectivity to the greatest extent and reduce computational complexity [9]. The computational complexity of GSConv is only about 50% of that of general convolution algorithms, but its learning performance has reached the level of general convolution algorithms. This method achieves lightweight model while ensuring model accuracy.

GSConv convolution is a convolution model that integrates conventional convolution (channel-enhanced convolution), depth-wise separable convolution, and channel permutation (Shuffle). As can be seen from Figure 7 , the deep separable convolution architecture used in this study innovatively introduces a channel rearrangement strategy, which can effectively combine the features generated by standard convolution with the features extracted by deep separable convolution. Through this feature integration mechanism, the model not only maintains the advantage of lightweight, but also significantly improves the feature expression ability. This method uses a uniform mixing method to effectively communicate local feature information between multiple channels , thereby achieving effective fusion of multiple features without any complex operations.

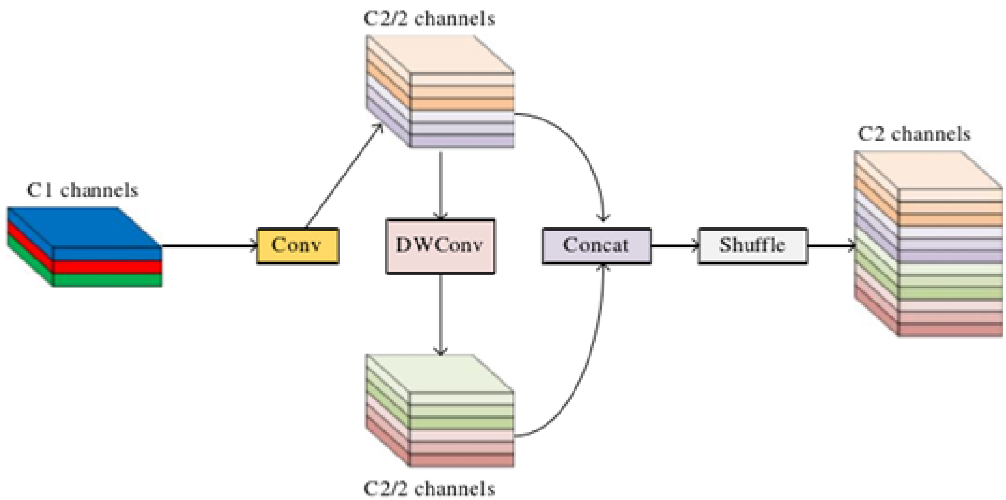

Figure 7 GSConv convolution structure diagram

In addition, based on the GSConv convolutional network, the residual network merging method is used to design the residual network GSCLayer structure to reduce the computational complexity and network structure complexity. Figure 8 shows a GSCLayer structure.

If the GSConv structure is used in all stages of YOLOv7, the network structure of the overall model will become more complex. At the same time, the deep network's resistance to data flow will also be enhanced, resulting in a significant increase in inference speed. But when these features are sent to the detection head, they have been stretched (maximum channel size and minimum width and height size). Based on this, this paper uses the GSConv method to replace the conventional convolution in the head area, and converts the global ELAN structure of the head area to the GSConv layer structure, thereby improving the detection speed of the entire area.

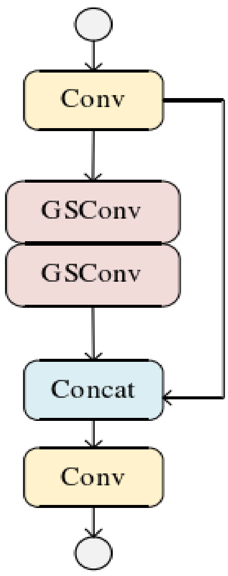

Figure 8 GSCLayer structure diagram

2.3 Adding SimAM attention mechanism

However, in water, light is affected by factors such as absorption and scattering, resulting in blurred images, which reduces the accuracy of the YOLOv7 model in detecting underwater targets. Therefore, this paper introduces an attention mechanism in object detection to increase the sensitivity of the object to the object, while maintaining the speed of the object's movement and improving the detection accuracy of the object.

At present, the attention mechanism of machine vision is mainly based on two aspects: channel and spatial domain. These two attention mechanisms are consistent with the "attribute attention" and "position" attention in the human brain. The existing attention mechanism has two main shortcomings: first, it can only finely process features in a specific channel or space, and cannot fully understand the changes of attention weights with channels and spaces [10]. Second, the structure of the model is composed of a series of complex factors, which will increase the number of parameters in the network. However, inspired by the brain's attention mechanism, SimAM's attention mechanism [11] can transform the attention model into a 3D weighted attention module without adding additional parameters.

Research shows that the SimAM attention mechanism can dynamically allocate neurons with important information and give priority to them. Its main features are: in underwater acoustic signal processing, it suppresses the spatial response of underwater targets to reduce the impact of complex background on target detection and improve target recognition accuracy. The calculation process is as follows:

$$\hat{X} = sigmoid\left(\frac{1}{E}\right) \otimes X$$

$$E = \frac{4(\sigma^2 + \lambda)}{(t - \mu)^2 + 2\sigma^2 + 2\lambda}$$

$$\mu = \frac{1}{Q}\sum_{i=1}^{Q} x_i$$

$$\sigma^2 = \frac{1}{Q}\sum_{i=1}^{Q} (x_i - \mu)^2$$

Where, $\hat{X}$is the target characteristic map after the underwater image is enhanced; where E is the energy function of each channel, the smaller its energy is, the more it is distinguished; to avoid E value being too large, it is constrained by sigmoid function; b is the dot product operation. Where, X is the input of the underwater object characteristic mapping; σ^2 is the input variance of each channel of the underwater target; where, t represents the neuron in the water; where, μ is the average of each channel in the underwater target characteristic map; ix is the other neuron that inputs the feature information into the i-th channel.

However, since the attention mechanism requires additional network parameters, we cannot introduce it into the network. As can be seen from Table 1 , Model 1 is the original YOLOv7 network, Model 2 adds the SimAM attention mechanism after each MPConv layer after the backbone network is modified, and Model 3 adds this mechanism after the first MPConv layer. Experiments show that the performance of Model 3 is improved by adding the SimAM attention mechanism after the first MPConv layer. Finally, in order to achieve a balance between performance and parameters in the MPConv layer.

Table 1 Comparison of adding attention mechanisms at different positions

Model	m AP
Model1	79.49%
Model2	81. 88%
Model 3	83.26%

In general, this paper proposes to use NewDSCLayer to replace the ELAN structure in the YOLOv7 framework, which can effectively solve the problem of reduced underwater target detection accuracy caused by the overlap of detected objects. The method of using GSCLayer to replace the head structure of the YOLOv7 structure detector can effectively solve the problem of poor real-time performance of the original YOLOv7 network; the SimAM attention mechanism is introduced into the YOLOv7 framework, which improves the accuracy of small target detection without adding redundant parameters; while taking into account the spatiotemporal correlation, the optimal energy function is constructed to explore the importance of each neuron and give a fast analytical solution to the functional function.

2.4 Add Grad-CAM visualization heat map framework

To address the above issues, this paper uses the Classification Activation Mapping (CAM) method to visualize the features of deep neural networks. Different from the traditional CAM algorithm, in terms of feature extraction, we abandon the traditional pooling strategy based on global average and use a fully connected layer to adjust the number of output nodes to match the number of target classes, thereby improving the pertinence of feature representation[12]. However, its disadvantage is that the entire network structure must be modified and the model must be retrained, which will take a lot of time and cost.

Grad-CAM is an improved version of CAM, which calculates the weights of each feature map by outputting the gradient relationship between the class score and the feature map. Grad-CAM has the same mathematical effect as CAM. However, the advantage of Grad-CAM is that it does not require modification of the network structure or reconstruction of the model.

The algorithm first finds the gradient relationship between the score of each category and the feature map of the last convolution layer. The gradient is used to calculate the importance of each feature map, thereby assigning weights to the feature maps. Finally, the weights of the feature maps are added together to obtain the final category activation map.First, the pixels of each feature map are averaged and the weights of each feature map are calculated. Finally, the weights of each feature map for each category are obtained. After finding the weight corresponding to each feature map, each feature map is weighted, and then all feature maps are added together to obtain the final visual heat map.

$$L^c_{Grad-CAM} = \mathrm{ReLU}(\sum_k \omega^c_k A^K)$$

Where AK is the kth feature map; ReLU is the activation function.

The Grad-CAM heatmap visualization architecture is used to analyze the attention distribution in the network. Under ideal conditions, the YOLOv7 network only focuses on four different types of underwater objects, including sea cucumbers, sea urchins, starfish, and scallops, without considering contextual information to ensure that the algorithm only considers the characteristics of the object itself without considering other interference.

The experiment shows that by training the original network, underwater targets can be well identified, which means that when detecting targets, more attention is paid to the area where the organisms are located, which shows that the model has a strong generalization ability. However, their attention is focused on the shadows under the rocks, which shows that the original network detection ability is not strong enough. For this reason, the Grad-CAM visual system is used to evaluate the underwater object detection method and make corresponding improvements accordingly.

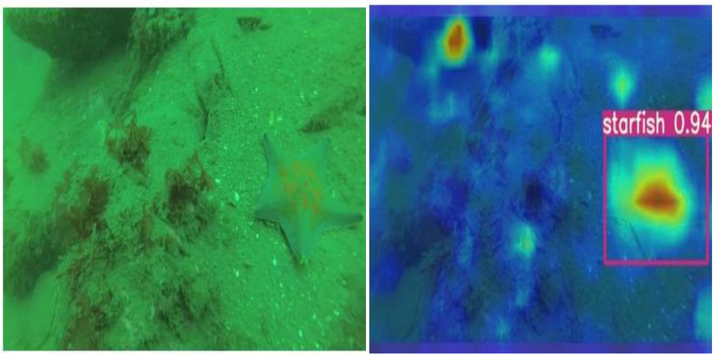

Figure 9 Comparison of Grad-CAM running results

3 Research on network security protection methods based on deep learning

3.1 Classification and identification of network security threats

Information security threats are complex and dynamic. Classifying and identifying them is a prerequisite for establishing effective defense measures. Threats include malware, botnets, phishing, denial of service, etc. Malicious programs are distributed in various ways such as email attachments and unsafe pages to achieve illegal access, data or normal system operation. "Botnet" refers to devices with malicious programs, which are used to launch large-scale network attacks. To effectively identify them, behavioral analysis and feature matching are required. Behavioral analysis refers to the detection of anomalies in networks and systems to discover possible malicious behaviors . For example, a malicious program may attempt to connect to a remote server to obtain instructions or change system settings to hide its existence. Feature matching technology is suitable for identifying known malicious code by identifying known threat code [13]. Deep learning is an important research direction in the field of threat identification. In addition, the real-time, accuracy and scalability of IDS are required to be met to a certain extent. Intrusion detection technology based on deep learning can automatically learn and identify complex attack patterns from massive amounts of data, and use this as a baseline to issue early warnings for behaviors that exceed the baseline, thus providing an effective way for intrusion detection.

3.2 Design and Optimization of Adaptive Defense Strategy

The core of adaptive defense strategy is to dynamically adjust to the dynamic changes in the network environment. When formulating strategies, we must fully consider flexible strategies and rapid response capabilities. Strategies should be able to automatically analyze network behavior, detect anomalies, and effectively handle them. Machine learning models represented by deep learning can be used to improve the intelligence of decision-making and the predictive ability of decision-making. By continuously learning the laws of network communication, potential threats can be predicted and identified to achieve the purpose of early warning. Model learning requires the use of high-quality and diverse data and can effectively identify different attack methods [14]. At the same time, the solution also has the ability to respond quickly and can respond to threats as soon as they are discovered, such as isolating infected devices and blocking malicious communications. In addition, in order to cope with the growing network environment and changing threats, the scalability and maintainability of policies are also very important. By dynamically monitoring and analyzing the network, the overall security protection level can be improved and the stability and security of the system can be ensured.

3.3 Adaptive Network Security Protection System Architecture

In order to ensure the efficiency, security and scalability of the system, a series of designs must be made in the architecture, including module design, hierarchical structure, real-time monitoring, automatic response, data-driven decision-making, and human-computer interaction.

During the integration process, the model needs to be seamlessly connected to the existing network security system and use middleware such as API for real-time data interaction. The automatic expansion process ensures the consistency and reliability of the model. Container technology. The constructed model is monitored in real time to ensure the accuracy, response time and resources occupied by the constructed model, so as to ensure the stability and efficiency of the constructed model. This feedback mechanism enables the model to automatically optimize the current network threats. In the integration process, security enhancement is an important link that cannot be ignored. The input and output of the model must be strictly checked to avoid security risks [15]. The user interface should be simple and intuitive, so that security analysts can monitor the status of the model and intervene when necessary. Finally, under the premise of meeting relevant legal provisions and protecting user privacy, the security and compliance of the data are guaranteed, so that it can be better applied to the adaptive network security protection system and improve its intelligence and automation. Table 2 lists the integration and configuration parameter values of the research model.

Table 2. Deep reinforcement learning model integration and deployment parameters

type	parameter	Deployment considerations
DQN	Learning rate, discount factor, exploration rate	Real-time data interaction, API integration, and automated deployment
Policy Gradient	Reward function design, policy network architecture	Stability, convergence speed, user interface design

3.4 Experimentation and evaluation of network security protection strategies

In order to achieve the protection of different types of network security, a network simulation platform consisting of 10 high-performance servers and 50 virtual machines was used to achieve the protection of different types of network security. The simulation platform simulated the multi-level complex network environment composed of intranet, DMZ and Internet access points in an all-round way. To ensure the authenticity and effectiveness of the experimental results, a large number of commonly used applications and services such as Web servers, mail servers, and databases were configured in the network. In the process of establishing the data set, we collected more than 1 TB of network traffic, including both normal traffic and attack traffic generated by our simulation technology. The database contains 100,000 normal behavior samples and 100,000 abnormal behavior samples, covering 20 attack modes and various security threats, such as DDoS attacks, SQL injection attacks, and cross-site scripting attacks. Through data enhancement, business perturbations, protocol changes and other methods, the promotion performance and robustness of the model are improved. The experimental results show that this method has strong adaptive ability and can effectively solve the modeling problem in a complex network environment.

Verifying its performance is a key step in evaluating its effectiveness and stability. Among them, detection accuracy and processing delay are key technologies that determine the efficiency and reliability of the protection system. The test results are detailed in Table 3.

Table 3 Experimental evaluation results

index	Target value	Current Value
Accuracy	> 9 3%	99.8 %
Accuracy	> 92%	96%
Response time	< 1 min	0.4min
System stability	99.9%	99.8 %
User interface usability	high	high

As shown in Table 3, the adaptive network security protection strategy has excellent performance in the main effectiveness metrics. The prediction accuracy is as high as 99.8%, which is 93% higher than the expected index, the experimental results show that the algorithm proposed in this paper has a good recognition effect on network attacks. The accuracy of this model is 96%, which is higher than 92%, indicating that this model also has an outstanding effect in reducing the false positive rate. The experiment proves that the algorithm proposed in this paper can respond quickly to emergencies in a relatively short time. The stability of the method reaches 99.8%, which is only slightly lower than the expected 99.9%, but still has high stability. The ease of use is "high", this system can well meet the needs of users and provide users with a good user experience. In short, this method has the characteristics of high efficiency, high precision, fast response and stability. fast response speed, good stability, etc., which can provide strong support for network security protection.

Conclusion

Based on the analysis of the underwater moving target detection method of YOLOv7, a new type of lightweight underwater moving target detection network ND-GS-YOLOv7 is designed in combination with the YOLOv7 network. On the one hand, the structure is improved: the ELAN structure in the original network is replaced by the NewDSCLayer structure, thereby improving the recognition accuracy of the target; the GSConv module was used to replace the original ELAN structure, which greatly improved the speed of model reasoning. On this basis, the SimAM attention model was combined to improve the network's ability to extract fine features and significantly improve the accuracy of small-scale object recognition. This method can improve the recognition ability of small targets while maintaining detection efficiency.The Grad-CAM visualization technology is used to deeply evaluate and optimize the network structure.

This study successfully built and verified an adaptive network security protection scheme based on deep reinforcement learning. The core of this scheme is to use the self-learning and adaptability of deep reinforcement learning to achieve accurate identification of network threats and dynamic optimization of defense strategies.

Experimental results show that this scheme performs well in the field of network security protection, can effectively identify a variety of network threats, and automatically adjust defense mechanisms according to threat characteristics and dynamic changes, thereby providing comprehensive network security protection.

References

[1] Ding Yuanming, Xu Lihua, Hou Mengke. Underwater target recognition technology based on deep learning[J]. Ship Science and Technology, 2024, 46(01): 143-147.

[2] Ma Minhui, Wang Hongru, Wang Jia. Underwater image enhancement algorithm based on improved MSRCR-CLAHE fusion[J]. Infrared Technology, 2023, 45(01): 23-32.

[3] Chen Yue, Luo Yihao, Li Jin. Application concept of underwater target detection based on multimodal perception[J]. Digital Ocean and Underwater Attack and Defense, 2024, 7(03): 334-341.

[4] Yang Xiaotian, Tan Jinlin, Yu Xin, et al. Ship target tracking in remote sensing images based on GAM-YOLOv8[J]. Journal of Jilin University (Earth Science Edition), 2025, 55(01): 328-339.

[5] Li Huakun, Zhao Lei, Li Heng, et al. Underwater image enhancement based on multi-scale fusion and detail enhancement[J]. Mechanical Science and Technology, 2023: 1-9.

[6] Wang CY, Bochkovskiy A, Liao HY M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[M]. arXiv, 2022: 1-10.

[7] Qian Xiaoqi, Liu Weifeng, Zhang Jing, et al. Degraded feature enhancement algorithm for underwater image target detection[J]. Journal of Image and Graphics, 2022, 27(11): 3185-3198.

[8] Zhuang Qiaohui. Research on multi-scale target detection and recognition methods in complex scenes based on deep learning[J]. Information Recording Materials, 2025, 26(02): 208-211.

[9] Liu Xiaochun, Yang Yunchuan, Hu Youfeng, et al. Data enhancement based on underwater moving target bright spot image model[J]. Journal of Northwestern Polytechnical University, 2024, 42(03): 417-425.

[10] Xie Fu, Zhu Dingju. A review of deep learning object detection methods[J]. Journal of Computer Systems and Applications, 2022, 31(2): 1-12.

[11] Huang Haining, Li Baoqi, Liu Jiyuan, et al. Review and Prospect of Underwater Target Recognition in Sonar Images[J]. Journal of Electronics & Information Technology, 2024, 46(05): 1742-1760.

[12] Han Tianbao, Wang Yue, Ren Shichang, et al. Underwater camouflage target detection method for unmanned surface vehicles[J]. Ship Science and Technology, 2024, 46(19): 85-91.

[13] Chen Jianrun, Mao Weining. Bearing-only underwater target maneuver detection based on deep learning[J]. Command Control & Simulation, 2024, 46(03):95-101.

[14] Li C, Li L, Jiang H, et al. YOLOv6: A single-stage object detection framework for industrial applications[J]. arXiv 2022, arXiv: 2209.02976.

[15] Wu Bin, Fang Zhen. Automatic recognition of underwater targets in side-scan sonar images[J]. Port Channel and Offshore Engineering, 2024, 61(02): 85-88+99.