International Journal of Multidisciplinary Research and Publications
ISSN (Online): 2581-6187

An Overview of Target Detection

YiMeng Zheng
School of Management Science and Engineering, Anhui University of Finance and Economics
Email address: 20212693@aufe.edu.cn

***Abstract*—*The field of object detection involves multiple aspects such as network structure, loss function, and evaluation methods. DCNN, YOLO, and SSD are common algorithms, and positional loss is a critical part. The COCO dataset is used to evaluate object detection, segmentation, and keypoint detection algorithms, with accuracy, recall, and F1 score being commonly used evaluation criteria. The future outlook includes improving algorithms and finding more effective evaluation methods.*

***Keywords*— *Target Detection, YOLO, DCNN, F1 score*

I. INTRODUCTION

Object detection [1] is a key task in the field of computer vision, aimed at accurately locating and identifying specific objects in images or videos. By using rectangular bounding boxes to locate detected objects and using classifiers for classification, this field is closely related to object classification, semantic segmentation, and instance segmentation. Object detection has broad application value in many fields [2], such as autonomous driving, security monitoring, medical image analysis, industrial production, and object tracking.

With the continuous advancement of technology, object detection algorithms have undergone an evolution [3] from traditional methods to deep learning. Deep neural networks [4] have made significant progress in recent years and have become the mainstream technology in the field of object detection. Through extensive data training and complex model design, deep learning can automatically extract features from images and perform classification [5], thereby achieving high-precision object detection.

This article will focus on exploring object detection techniques based on deep neural networks. We will delve into the application of deep learning in object detection, including its principles, advantages, and challenges.

II. NETWORK STRUCTURE

In the field of computer vision, network architecture is a key element in achieving image processing and analysis. With the rapid development of deep learning technology, the design and optimization of network structures have played a crucial role in visual tasks. The following paragraph will introduce several commonly used network [6] structures at present.

A. *Deep Convolutional Neural Network (DCNN)*

Deep Convolutional Neural Network (DCNN) is a deep learning architecture specifically designed for processing visual data [7]. Its emergence marks a significant leap in the field of computer vision, especially in tasks such as image recognition, object detection, and semantic segmentation.

The core idea of DCNN is to extract advanced features of the input image through multi-level convolution and pooling operations [8], thereby achieving automatic learning of objects and patterns in the image. The structure of a network usually includes convolutional layers, pooling layers, and fully connected layers. The convolutional layer is responsible for extracting spatial features of the image, the pooling layer is used to reduce the dimensionality of the data and enhance the translation invariance of the network, and the fully connected layer is responsible for mapping the extracted features to the output category.

The success of DCNN is attributed to its training ability on large-scale datasets and automatic learning ability on complex features. By continuously adjusting network parameters through backpropagation algorithms, DCNN can gradually optimize itself to meet the requirements of specific tasks. Famous DCNN architectures include LeNet, AlexNet, VGG, GoogLeNet, and ResNet, which demonstrate excellent performance on different tasks.

At present, DCNN is widely used in fields such as image classification, object detection, facial recognition, and medical image analysis. Due to its ability to automatically extract and learn complex features from images, DCNN has strong adaptability for processing large-scale and high-dimensional data, making it an indispensable tool in the field of computer vision.

B. *Regional Convolutional Neural Network (RCNN)*

Regional Convolutional Neural Network [9] is an object detection framework based on Convolutional Neural Network (CNN), aimed at overcoming the challenges of traditional object detection algorithms in handling object localization and classification tasks. The design idea of RCNN is to introduce a Region Proposal Network (RPN) to select candidate regions in the image that may contain targets, and to perform deep learning on these regions through CNN.

The workflow of RCNN typically includes four main steps. Firstly, a large number of candidate regions are generated through methods such as selective search, which are considered to potentially contain targets. Secondly, each candidate region is propagated forward through CNN to extract features and map them to a high-dimensional space. Subsequently, these features are input into Support Vector Machine (SVM) for target classification. Finally, the bounding boxes of the candidate regions are fine tuned using a regressor to further improve the accuracy of target localization.

The innovation of RCNN lies in the introduction of RPN, which can learn to generate candidate regions end-to-end, resulting in significant performance improvement of the entire system in object detection tasks. Some variants of the RCNN framework include Fast R-CNN [10] and Fast R-CNN, which further accelerate the object detection process and improve real-time performance.

Although the RCNN series models have achieved success in the field of object detection, their high computational complexity leads to relatively slow training and inference speeds. Over time, subsequent models such as YOLO (You Only Look Once) and SSD (Single Shot Multibox Detector) have gradually emerged, providing more efficient object detection solutions. However, RCNN still plays an important driving role in leading the development of the field of object detection.

C. You Only Look Once （YOLO）

You Only Look Once [11] is an advanced target detection algorithm, which is famous for its efficiency and real-time performance. Different from the traditional target detection method, YOLO uses a single neural network model to complete the target location and classification at the same time through a forward propagation. This unique design makes YOLO a highly effective real-time target detection solution.

The core idea of YOLO is to divide the image into grids and predict the location of the bounding box and the category of objects in each grid cell. This mesh generation method allows YOLO to obtain global information on a complete image at one time, so that it can capture the global context of the object without being limited by the local information of the traditional sliding window method.

The output of the algorithm is a list of all detection boxes, each box with its coordinates, category and confidence score. Due to its high efficiency, YOLO has been widely used in real-time video analysis, automatic driving, security monitoring and other fields.

The development of YOLO has gone through several versions, including YOLOv1, YOLOv2 (also known as YOLO9000), YOLOv3 and YOLOv4. Each version has improved its algorithm performance, speed and diversity, making YOLO series an advanced technology that has attracted much attention in the field of target detection. Its simple and efficient design makes YOLO an ideal choice for processing large-scale real-time target detection tasks, and promotes the rapid development of target detection in the field of computer vision.

D. Single Shot Multibox Detector （SSD）

Single Shot Multibox Detector (SSD) [12] is an advanced target detection algorithm, which aims to achieve efficient and accurate real-time target detection. The design concept of SSD is to complete multi-scale detection of targets through a single forward propagation, so that it has strong adaptability and efficiency.

One of the key innovations of SSD is the introduction of multi-scale feature maps, which can effectively capture the scale and shape information of various targets through target detection on feature maps at different levels. This enables SSD to maintain good performance when dealing with small targets and large targets, and has wider applicability.

By performing classification and bounding box regression on predefined anchors on each feature map, the algorithm can simultaneously detect targets with multiple scales and aspect ratios. The output of SSD includes the category probability and location information of each detection frame, and the final detection results are screened through post-processing steps such as non maximum suppression (NMS).

SSD has achieved remarkable success in the field of target detection. Its fast detection speed and high accuracy make it an ideal choice for many real-time application scenarios, such as video analysis, automatic driving, intelligent monitoring, etc. The continuous evolution of algorithms has also led to multiple SSD versions, such as SSD300, SSD512, etc., to meet the requirements of different tasks and resource requirements. In general, SSD, as an advanced technology of multi-scale target detection, is of great significance for improving the balance between real-time performance and accuracy.

III. LOSS FUNCTION

In the target detection task, the loss function is the key factor to evaluate the performance of the mode [13], which usually consists of two main components: location loss and classification loss. These two loss functions are respectively responsible for measuring the prediction accuracy of the model for target location and target category in the training process, so as to guide the model to learn how to accurately locate and correctly classify and detect targets.

A. Localization Loss

Position loss focuses on the accuracy of the measurement model in predicting the location of the target bounding box. A common position loss is Smooth L1 Loss, and its formula is as follows:

$$\begin{cases} 0.5(x-\hat{x})^2 & if|x-\hat{x}|<1 \\ |x-\hat{x}|-0.5 & \textbf{otherwise} \end{cases}$$

Where, x is the predicted boundary box position of the model, $\hat{x}$ is the boundary box location of the actual dimension.

B. Classified Loss

Classification loss measures the accuracy of the model in classifying target categories. Common classification losses include cross entropy loss:

$$L_{cls}(p,\hat{\mathbf{p}})=-\sum p_i\hat{\mathbf{p}}_i\cdot\log(p_i\hat{\mathbf{p}})$$

The total loss function is usually composed of the combination of location loss and classification loss. For example, the total loss in Faster R-CNN can be expressed as:

$$L = L_{loc}+\lambda L_{cls}$$

Among them, λ Is a weight parameter used to balance the influence of both. This comprehensive loss function is designed to enable the model to accurately locate and classify targets in training, so as to improve the overall performance of the target detection system.

YiMeng Zheng, “An overview of target detection,” *International Journal of Multidisciplinary Research and Publications (IJMRAP)*, Volume 6, Issue 8, pp. 158-163, 2024.

IV. COMPLEX SITUATION OF TARGET DETECTION

Target detection faces many complex situations, some of which mainly include:

A. Target occlusion

Part or all of the target in the image is occluded by other objects, scene elements or occlusions, which is called target occlusion. This situation may be caused by a variety of factors, including occlusion of foreground objects, background elements, or other adjacent objects, which will hinder the visual information of the object, making the model face challenges in locating and classifying occluded objects.

In daily scenes, target occlusion [14] is a common problem. For example, in traffic monitoring, cars may be blocked by other vehicles, buildings or road signs; In the crowd, individuals may be blocked by other people, objects or structures; In natural scenes, animals or plants may be blocked by trees, rocks or other obstacles.

To solve the problem of target occlusion, researchers have proposed some innovative solutions:

Multi scale detection: Multi scale detection method is used to enable the model to detect objects of different sizes, so as to better deal with occlusion problems. For example, convolutional neural network (CNN) with pyramid structure is used to process information of different scales.

Occlusion awareness model: introduce the design of occlusion awareness, so that the model can better understand and process the occluded target in the learning process. This may involve special processing of occluded parts and speculation of occluded areas.

Generate countermeasure network (GAN): use the generation ability of the generated countermeasure network to help the model better understand the shape and structure of the occluded target by generating a possible target appearance from the occluded area.

Occluded area completion: using image repair technology, try to complete the image by inferring the content of the occluded part, so as to provide more complete information to the target detection model.

Scale change: the size of the target in the image may change due to the change of distance, angle or camera angle of view. The model needs to have the ability to effectively detect and recognize targets of different scales.

Perspective change: the target presents different appearances under different perspectives, which may cause the model to be difficult to correctly identify the target under certain perspectives, especially when the training data does not cover all possible perspective changes.

B. Category Imbalance:

In the target detection task, the unbalanced number of targets of different categories in the dataset [15] means that the number of samples of each target category is significantly different. This means that some categories have more samples, while others have less samples. This imbalance may lead to the model being more prone to learning the categories with higher frequency in the training process, while performing poorly for the rare categories.

In actual scenarios, the problem of category imbalance in data sets often occurs. For example, in traffic monitoring, cars may be the main target, while rare traffic signs or non motor vehicles may be relatively few. In medical images, normal samples may be far more than abnormal samples. This imbalance may also occur in natural scenes, such as in animal recognition, where the number of samples of some common animals may be far more than that of some rare species.

In order to solve the problem of category imbalance, researchers have proposed a variety of methods:

Resampling: This includes means such as oversampling rare categories and undersampling common categories to make the number of samples in different categories more balanced.

Class Weighting: by assigning different weights to samples of different categories, the model pays more attention to learning rare categories in training.

Online Hard Example Mining: Dynamically adjust the loss function to pay more attention to difficult samples in training, which may include some rare categories.

Ensemble Learning: combine the prediction results of multiple models to improve the detection performance of rare categories.

Generative Approaches: generate additional rare category samples by generating countermeasures networks (GAN) and other methods to expand the dataset.

C. Deformation and distortion[16]

The shape change of the target usually means that the target may appear different in the image due to deformation or distortion. This situation increases the difficulty of target detection, because the model needs to have the ability to accurately detect targets with different shapes and attitudes. For example, cars may show a variety of appearances in different forms, such as different angles, deformation states or occlusion, while animals may also show different shapes in different poses.

In the actual scene, it is very common for the target shape to change. In traffic monitoring, vehicles may experience lane changing, turning and other operations, resulting in changes in their shape. In medical images, organs may present different shapes due to different positions or deformations. In dynamic scenes, the change of human posture may also pose a challenge to target detection.

In order to solve the difficulties caused by target shape changes, researchers have proposed some methods:

Pose Estimation: use the attitude estimation technology to predict the attitude information of the target, so that the model can better understand the specific shape of the target in the image.

Multi scale feature extraction: use multi-scale feature extraction methods to ensure that the model can capture the shape information of the target at different scales.

Shape Models: use prior shape models to model the shape of the target, so as to help the model better understand the changes of the target under different shapes.

Data enhancement: introduce shape changing samples into the training data, and transform the original image through

rotation, distortion, etc. to increase the robustness of the model to the diversity of target shapes.
Reinforcement Learning: use reinforcement learning methods to enable models to learn adaptively according to different shape changes.

D. Small target detection[11]

Small size objects in the image refer to those objects that occupy fewer pixels in the image. Due to their relatively small physical size, the model faces a series of challenges in detecting these objects. In this case, the model needs to accurately locate and classify targets with only limited pixel information, which increases the complexity of the detection task.

In practical scenarios, the presence of small-sized targets is quite common. For example, in aerial images, distant buildings, vehicles, or people may appear smaller in size due to their distance. In medical images, small lesions or cells may require high-precision detection. In surveillance videos, small-sized targets may include distant people, vehicles, or other objects.

To solve the problem of detecting small-sized objects, researchers have adopted various strategies and techniques:
Multi scale detection: Introducing multi-scale detection methods to capture feature information of small-sized targets by detecting targets at different scales.
Pyramid network structure: Using a pyramid network structure, construct multi-level feature pyramids to ensure that the model can extract features suitable for small object detection from different levels.
Attention mechanism: Introducing attention mechanism enables the model to pay more attention to small-sized target areas in the image, improving its perception ability towards small targets.
Data augmentation: Introduce data augmentation methods such as scaling, rotation, and random cropping for small targets in training data to better adapt the model to the diversity of small targets.
High resolution image: Provides higher resolution input images to increase the visible information of small-sized targets, thereby improving the detection performance of the model.

E. Dynamic object detection

Dynamic object detection refers to the detection and tracking of target objects in video sequences, where their position, appearance, or state may change over time. This situation increases the complexity of object detection, as the model needs to continuously locate and classify targets in different frames of the video, taking into account the possible motion, deformation, and interaction with other targets.

In practical applications, dynamic object detection is widely used in multiple fields:
Autonomous driving: In the field of autonomous driving, dynamic object detection is used to identify and track other road users, such as pedestrians, vehicles, and bicycles, to ensure that vehicles can travel safely and make appropriate decisions.
Video surveillance: In monitoring systems, dynamic object detection can be used to monitor moving objects in real-time, providing sensitivity to security and abnormal events.
Medical imaging: In the medical field, dynamic object detection can be used to track moving organs, cells, or injection objects, which is of great value for surgical navigation and real-time image guidance.
The methods for solving dynamic object detection problems include:
Optical flow estimation: Using optical flow estimation methods to capture the motion information of targets between consecutive frames, to help locate and track dynamic targets.
Convolutional Neural Network (CNN): Utilizing the convolutional neural network structure in deep learning, it improves the detection accuracy of moving targets by learning the features of the targets.
Multi target tracking algorithm: Using multi-target tracking algorithms, such as Kalman filters or correlation filters, to establish the trajectory of targets between consecutive frames.
Spatiotemporal modeling: Introducing the method of spatiotemporal modeling to model the motion and changes of targets in time series, in order to improve the understanding and prediction ability of dynamic targets.
Transfer learning: Utilizing the idea of transfer learning, fine-tuning the model trained in static scenes in dynamic scenes to improve its generalization performance.

Zhang [17] proposes a lightweight object detection method based on optimized RetinaNet model feature fusion for dynamic object detection. By using depth wise separable convolution and spatial feature fusion mechanisms, the model parameter count is reduced and feature scale invariance is improved. Meanwhile, a feature adaptive fusion target tracking algorithm based on normalized attention mechanism was proposed, which suppresses weights through lightweight attention mechanism and enhances the last four feature layers of the backbone network through path enhancement. The experimental results show that the optimized object detection algorithm greatly improves the inference speed of the network while ensuring accuracy, while the proposed tracking algorithm achieves a high success rate in the constructed dataset and has strong robustness. Finally, a target detection and tracking system was designed to test the target tracking system in practical scenarios, achieving good tracking results, proving the applicability of the optimized algorithm and design system.

V. DATASETS AND EVALUATION METHODS

A. Common Dataset References

COCO (Common Objects in COntext):
Features: Large scale, strong diversity, including 80 different categories of goals.
Usage: Widely used for evaluating object detection, segmentation, and keypoint detection algorithms.
PASCAL VOC (Visual Object Classes):
Features: Contains 20 categories of targets, as well as corresponding images, annotations, and training/test set partitioning.
Usage: As one of the classic benchmark datasets for object detection tasks, widely used for algorithm evaluation.

ImageNet:
Features: Large scale, multi category, but more widely used for image classification. Contains the corresponding detection task dataset (ImageNet Detection).
Purpose: Used for object detection algorithm evaluation on the ImageNet Detection subset.
KITTI Vision Benchmark Suite:
Features: Targeted at autonomous driving scenarios, including stereo vision, LiDAR, and camera data for targets such as vehicles and pedestrians.
Purpose: Used for evaluating object detection, 3D object detection, and other visual tasks related to autonomous driving.
Open Images Dataset:
Features: Large scale, strong diversity, containing millions of images, each image may contain multiple targets.
Purpose: To provide large-scale data for tasks such as object detection and segmentation.
MSCOCO Keypoints Challenge:
Feature: Based on the COCO dataset, it provides annotation information for human key points, including human posture.
Purpose: Used to evaluate human keypoint detection and pose estimation algorithms.
Cityscapes:
Features: Targeted at urban scenes, including high-resolution images, annotated with different categories of traffic signs, vehicles, pedestrians, and other targets.
Purpose: Mainly used for object detection and segmentation algorithm evaluation in urban scenes.
ADE20K:
Features: Scene oriented understanding, including 20000 images, annotated with targets, objects, and scene elements in various scenes.
Purpose: Mainly used for research on scene understanding, semantic segmentation, and object detection.

These datasets play a crucial role in different application fields and research directions, providing important benchmarks and experimental platforms for the development of object detection algorithms. Choosing a dataset suitable for specific tasks and scenarios is crucial for research and algorithm development.

B. Common evaluation methods

Accuracy and recall rate:

Precision and Recall are commonly used evaluation metrics in object detection tasks, providing us with a way to gain a deeper understanding of model performance. These two indicators are calculated based on the number of true positive, false positive, and false negative examples in the model, using the following method:

Precision= True Positive/True Positive+False Positive

Among them, True Positive represents the number of targets correctly detected by the model, and False Positive represents the number of targets incorrectly marked by the model as non targets. The higher the accuracy, the less the model incorrectly marks non targets as targets in its predictions.

Recall:

The recall rate represents the proportion of successfully detected positive cases by the model relative to the actual positive cases, that is, the proportion of correctly recognized targets by the model to the total number of actual targets. The calculation formula for recall rate is:

Recall= True Positive/True Positive+False Negative

Among them, True Positive represents the number of targets correctly detected by the model, and False Negative represents the number of targets that the model failed to detect correctly. The higher the recall rate, the better the model can capture the real target.

F1 score:

F1 score is a commonly used evaluation metric in object detection tasks, which strikes a balance between precision and recall, providing a comprehensive evaluation of model performance.

Among them, Precision represents the proportion of true positive cases in the model's detection results, and Recall represents the proportion of positive cases successfully detected by the model relative to the true positive cases. The range of F1 scores is between 0 and 1, and the closer it is to 1, the better the balance between accuracy and recall of the model.

Usage: The F1 score is particularly suitable for dealing with imbalanced categories, where there is a significant difference in the number of positive and negative examples. In this case, using only precision or recall may not fully evaluate the performance of the model. The F1 score provides an effective evaluation method for achieving a balance between accuracy and recall while considering both factors comprehensively. In object detection tasks, the F1 score is usually used to evaluate the overall performance of the model in positive example categories, ensuring that the model achieves balanced results in detecting and identifying targets.

Taking into account the F1 score, accuracy, and recall, we can gain a more comprehensive understanding of the performance of object detection models in various aspects, providing powerful references for selecting appropriate models or adjusting thresholds.

VI. Conclusions

Object detection plays a crucial role in the field of computer vision, with the core task of accurately locating and identifying specific objects in images or videos. With the continuous development of technology, object detection technology has shown broad application prospects in scientific research and practical industrial applications.

The application fields of object detection cover multiple fields such as autonomous driving, security monitoring, medical image analysis, industrial production, and object tracking, providing key technical support for these fields. With the continuous advancement of technology, object detection algorithms are also evolving, gradually transitioning from traditional methods to deep learning. The rise of deep neural networks marks significant progress in the field of object detection, but its development is limited by computing resources, datasets, and theoretical foundations.

Although deep learning has achieved significant success in object detection, traditional object detection algorithms are

YiMeng Zheng, “An overview of target detection,” *International Journal of Multidisciplinary Research and Publications (IJMRAP)*, Volume 6, Issue 8, pp. 158-163, 2024.

still popular, especially in situations where computing resources are limited and data is limited, demonstrating stability and efficiency. Traditional algorithms such as Haar features and cascaded classifiers, Histogram of Oriented Gradients (HOG), Edge Direction Histograms (EDH), Color Histograms, and Background Modeling still have unique advantages in specific scenarios.

With the further enhancement of computing resources, the expansion of datasets, and the in-depth research of deep learning theory, object detection technology is expected to play a greater role in a wider range of application scenarios. Meanwhile, for specific scenarios and resource constrained environments, traditional object detection algorithms will continue to leverage their advantages of stability and efficiency. In the future, the field of object detection will achieve more comprehensive and innovative development in the fusion of deep learning and traditional methods.

REFERENCES

[1] Yin Hongpeng, Chen Bo, Chai Yi, et al. Review of Visual Based Object Detection and Tracking [J]. Journal of Automation, 2016,42 (10): 1466-1489. DOI: 10.16383/j.aas.2016.c150823
[2] Zhou Xiaoyan, Wang Ke, Li Lingyan. Overview of deep learning based object detection algorithms [J]. Electronic Measurement Technology, 2017,40 (11): 89-93
[3] Zhou Feiyan, Jin Linpeng, Dong Jun. Research Review of Convolutional Neural Networks [J]. Journal of Computer Science, 2017,40 (06): 1229-1251
[4] Sun Zhijun, Xue Lei, Xu Yangming, et al. A Review of Deep Learning Research [J]. Computer Application Research, 2012,29 (08): 2806-2810
[5] Liu Li, Kuang Gangyao. Overview of Image Texture Feature Extraction Methods [J]. Chinese Journal of Image and Graphics, 2009,14 (04): 622-635
[6] Lu Hongtao, Zhang Qinchuan. Review of Application Research of Deep Convolutional Neural Networks in Computer Vision [J]. Data Acquisition and Processing, 2016,31 (01): 1-17. DOI: 10.16337/j.1004-9037.2016.01.01
[7] Sun Zhijun, Xue Lei, Xu Yangming, et al. A Review of Deep Learning Research [J]. Computer Application Research, 2012,29 (08): 2806-2810
[8] Liu Wanjun, Liang Xuejian, Qu Haicheng. Study on the Learning Performance of Convolutional Neural Networks with Different Pooling Models [J]. Chinese Journal of Image and Graphics, 2016,21 (09): 1178-1190
[9] Zhou Junyu, Zhao Yanming. Review of Convolutional Neural Networks in Image Classification and Object Detection Applications [J]. Computer Engineering and Applications, 2017,53 (13): 34-41
[10] Wang Lin, Zhang Hehe The Application of Faster R-CNN Model in Vehicle Detection [J]. Computer Applications, 2018,38 (03): 666-670
[11] Ju Muran, Luo Haibo, Wang Zhongbo, et al. Improved YOLO V3 algorithm and its application in small object detection [J]. Journal of Optics, 2019, 39 (07): 253-260
[12] Zhang Shun, Gong Yihong, Wang Jinjun. Development of Deep Convolutional Neural Networks and Their Applications in the Field of Computer Vision [J]. Journal of Computer Science, 2019, 42 (03): 453-482
[13] Chang Feiliang, Ma Li, Qiao Yizheng. Target tracking algorithm based on feature correlation matching under occlusion [J]. Chinese Journal of Image and Graphics, 2006, (06): 877-882
[14] Xue Chen, Zhu Ming, Liu Chunxiang. Overview of Target Tracking Algorithms under Occlusion [J]. Chinese Journal of Optics and Applied Optics, 2009,2 (05): 388-394
[15] Song Guanghui. Research on Image Annotation Methods Based on Transfer Learning and Deep Convolutional Features [D]. Zhejiang University, 2017
[16] Song Lili, Wu Yadong, Sun Bo. An improved method for correcting document image distortion [J]. Computer Engineering, 2011, 37 (01): 204-206
[17] Zhang Yupeng. Research on Model Lightweight Target Tracking Algorithm [D]. Yanshan University, 2023. DOI: 10.27440/d.cnki. gysdu.2023.000576