Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 12, No. 3, 2024

Complex Scene Understanding and Object Detection Algorithm Assisted by Artificial Intelligence

Binrong Zhu[1, a], Guiran Liu[1, b]

[1]San Francisco State University, College of Science & Engineering (CoSE), San Francisco, United States
[a]bzhu2@sfsu.edu, [b]gliu@sfsu.edu

Abstract: The purpose of this paper is to study the algorithm of object detection and scene understanding in complex scenes assisted by AI(Artificial Intelligence), so as to improve the machine's perception of the surrounding environment. Aiming at occlusion and dense scenes, this paper proposes a series of innovative algorithms by introducing technologies such as multi-scale detection, occlusion processing and real-time optimization. Experimental results show that the proposed algorithm has achieved excellent performance on several public data sets. Compared with the current mainstream object detection algorithm - YOLO, the method proposed in this paper has a significant improvement in accuracy. Especially in multi-scale object detection, occlusion and dense scene processing and real-time optimization, the proposed algorithm shows obvious advantages. The research results provide an effective solution for object detection and scene understanding in complex scenes, and promote the application of AI system in related fields.

Keywords: Complex scene understanding, Object detection, Multi-scale detection, Occlusion treatment, Real-time optimization.

1. Introduction

In today's era of rapid information development, AI has become an important force to promote social progress and technological innovation [1]. Among them, complex scene understanding and object detection, as the core tasks in the field of AI vision, play a vital role in improving the machine's perception of the surrounding environment and realizing intelligent decision-making [2-3]. With the continuous progress of computer vision technology, especially with the promotion of deep learning algorithm, the ability of object detection and recognition in complex scenes has been significantly improved, which has brought revolutionary changes to many fields such as autonomous driving, intelligent security, robot navigation, medical image analysis [4].

Understanding complex scenes is not only a simple recognition of objects in images, but also a grasp of the overall semantics of the scene, such as identifying the elements such as roads, buildings and vegetation in the scene and their spatial relationships [5]. Object detection is to accurately locate and identify the target objects in the image, such as pedestrians, vehicles and animals, which is very important to ensure that the system can accurately respond to the key information in the surrounding environment [6-7]. Therefore, in-depth study of complex scene understanding and object detection algorithms is of far-reaching significance for improving the intelligence level of AI system and broadening its application scenarios.

2. Basic Theory of Complex Scene Understanding and Object Detection

2.1. Basic concepts of scene understanding

Scene understanding is an important branch of computer vision, which aims to make computers understand the scene content in images or videos like humans [8]. This includes identifying objects, backgrounds, spatial layouts and semantic relationships among them. Scene understanding involves not only low-level image processing techniques, such as edge detection and image segmentation, but also high-level semantic analysis, such as object recognition and scene classification. Through scene understanding, the computer can better explain the image content and provide support for subsequent decision-making.

2.2. Object detection basis

Object detection is a basic task in computer vision, which aims to accurately identify and locate the position and category of target objects from images or videos. Traditional object detection methods are mainly based on manually designed features and classifiers, such as HOG features combined with SVM classifier. However, with the rise of deep learning, especially the successful application of CNN in the field of image processing, the object detection algorithm has made remarkable progress. At present, mainstream object detection algorithms, such as YOLO and Faster R-CNN, are all based on deep learning framework, and realize efficient and accurate detection of target objects through end-to-end learning.

2.3. Feature extraction and representation learning

Feature extraction is a key link in object detection and scene understanding. Effective features can capture the key information in the image and improve the recognition performance of the algorithm. In the era of deep learning, feature extraction and representation learning are usually completed automatically through deep learning models such as convolutional neural networks. These models can learn rich feature representations from a large number of data, including low-level features such as shape, texture and color, as well as more advanced semantic features. Through feature extraction and representation learning, the algorithm can better understand the image content and improve the accuracy of object detection and scene understanding.

3. Object Detection Algorithm in Complex Scene

3.1. Multi-scale object detection

In complex scenes, the size and shape of objects often vary, from small parts to large buildings, which may appear in the same image. Multi-scale object detection aims to solve this challenge and ensure that the algorithm can accurately identify objects of different sizes. In this study, FPN(Feature Pyramid Networks), multi-scale training and testing strategies and attention mechanism in deep learning are used to enhance the detection ability of the algorithm for multi-scale objects. FPN is one of the core components of this study. By constructing a top-down path and horizontal connection, it combines high-level semantic information with low-level detail information to generate a feature pyramid with rich levels. This multi-level feature representation enables the algorithm to capture the fine features of small-scale objects and the global structure of large-scale objects at the same time, thus significantly improving the detection accuracy of objects of different scales. At the same time, the multi-scale training and testing strategy further enhances the generalization ability of the algorithm. In the training stage, we randomly adjust the size of the input image, so that the model can learn the object characteristics at different scales. This data enhancement method not only increases the diversity of training data, but also forces the model to extract and classify features effectively at different scales. In the testing stage, this paper also uses multi-scale input, adjusts the image to different sizes for multiple tests, and synthesizes these results to get more accurate final output. This strategy effectively reduces the missed detection and false detection caused by scale change, and improves the robustness of the algorithm.

The introduction of attention mechanism further optimizes the focusing ability of the algorithm on key information. The attention mechanism can be expressed by the following formula:

$$\alpha = \mathrm{softmax}\left(\mathrm{W}_{\alpha} \cdot \mathrm{ReLU}\left(\mathrm{W}_{\mathrm{att}} \cdot X\right)\right) \tag{1}$$

$$Y = \alpha \cdot X \tag{2}$$

Where X is the input feature map, $\mathrm{W}_{\mathrm{att}}$ and W_{α} are the learnable weights, ReLU is the activation function, α is the attention weight, and Y is the attention-weighted feature map.

In complex scenes, background noise and object occlusion often interfere with detection. By introducing attention mechanism, the algorithm can automatically learn and highlight those feature regions that are most critical to the detection task, while suppressing irrelevant background information. This mechanism not only improves the effectiveness of feature representation, but also makes the algorithm more robust when dealing with complex scenes.

3.2. Occlusion and dense scene processing

Occlusion and dense scene are two difficult problems in object detection. In the case of occlusion, some or all areas of an object may be occluded by other objects, which makes the detection more difficult. However, in dense scenes, multiple objects are closely arranged, which is easy to cause false detection and missed detection. In this study, context information, component model and other technologies are used to improve the performance of object detection in occluded and dense scenes. Using context information is an important means to improve the performance of object detection in occluded and dense scenes. In complex visual scenes, objects usually do not exist in isolation, and they are closely related to the surrounding environment and other objects. By capturing and analyzing these contextual information, the algorithm can better understand the overall structure of the scene, so as to locate the occluded or densely arranged objects more accurately. For example, when a person is detected, the algorithm can use the context information to infer the items that the person may carry (such as backpacks, handbags, etc.), thus improving the detection accuracy of these small objects.

Building component model is another effective strategy, especially suitable for dealing with partially occluded objects. The component model is expressed in the following ways:

$$O = \left\{\mathrm{P}_1, \mathrm{P}_2, \mathrm{P}_3, \ldots, \mathrm{P}_{\mathrm{n}}\right\} \tag{3}$$

$$P\left(\mathrm{P}_{\mathrm{i}} \middle| I\right) = \mathrm{PartDetector}_{\mathrm{i}}\left(I\right) \tag{4}$$

$$P\left(\mathrm{O} \middle| I\right) = \prod_{i=1}^{n} P\left(P_{\mathrm{i}} \middle| I\right) \tag{5}$$

Where O is an object composed of multiple components P_{i}, $\mathrm{PartDetector}_{\mathrm{i}}$ is a detector trained for the i component, $P\left(P_{\mathrm{i}} \middle| I\right)$ is the detection probability of the component P_{i} under the given image I, and $P\left(O \middle| I\right)$ is the detection probability of the object O under the given image I.

In this paper, the object is decomposed into several parts and an independent detector is trained for each part. In the detection process, even if some parts of the object are blocked, the algorithm can still infer the existence and position of the object by identifying other visible parts. This method not only improves the robustness of detection, but also enables the algorithm to describe the structural characteristics of objects more accurately.

In addition, this paper optimizes the computational efficiency and memory occupation of object detection algorithm by model pruning, quantization and lightweight network design. At the same time, parallel computing and hardware acceleration are used to further improve the real-time performance of the algorithm. Through these optimization measures, we can ensure the detection accuracy and achieve faster detection speed to meet the needs of practical application.

3.3. Experiment and result analysis

In order to comprehensively and deeply verify the effectiveness of the proposed algorithm in practical application, this section carefully designs and implements comparative experiments. This verification process aims to show the performance of the new algorithm in the object detection task through objective data and intuitive

visualization results, and compare it with the technical standards widely recognized by the industry, thus highlighting its innovative value and potential improvement space.

In order to show the performance level of the proposed algorithm more intuitively, we choose —YOLO, the current mainstream object detection algorithm, as the comparison object. YOLO occupies an important position in the field of object detection because of its high speed and high precision. In this paper, the proposed algorithm and YOLO algorithm are run on the same public data set, and their recognition accuracy is recorded. By comparing their performances on the same data set, we can see the advantages or disadvantages of the new algorithm more clearly. The results are shown in Figure 1 and Figure 2:

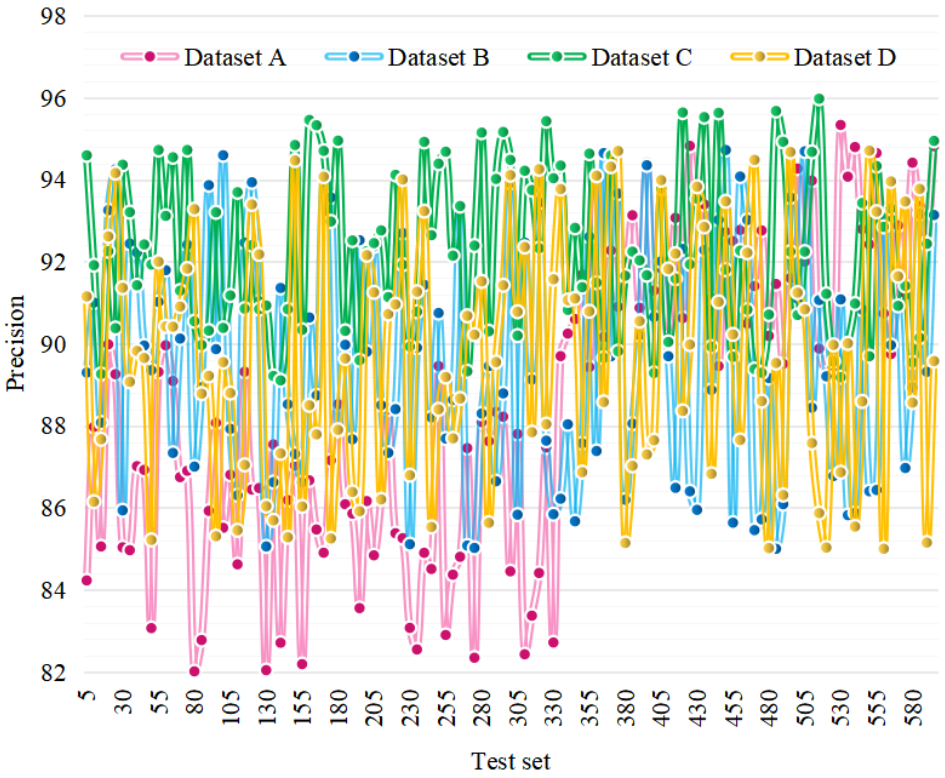

Figure 1. Recognition accuracy (Proposed algorithm)

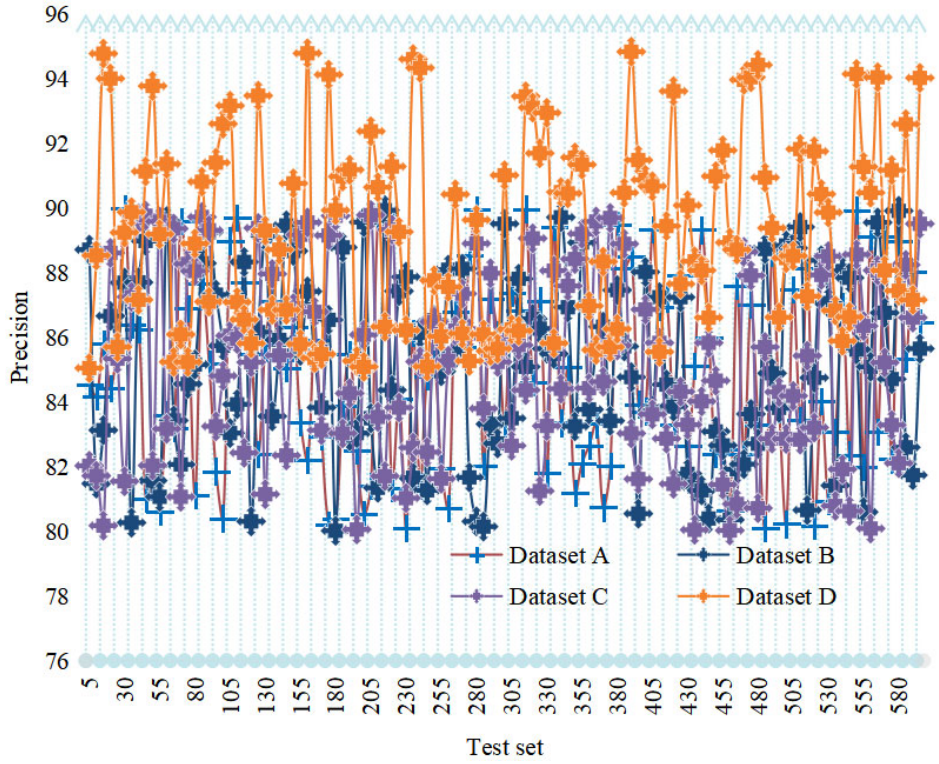

Figure 2. Recognition accuracy (YOLO algorithm)

It can be seen that the overall recognition accuracy of the proposed algorithm is 4.5% higher than that of YOLO (89.5% VS 85%), which shows that the proposed algorithm has certain advantages in overall detection ability.

In addition to the recognition accuracy, the calculation efficiency and memory occupation of the algorithm are also important indicators to measure its practical application value. Therefore, this paper further tests the performance of the proposed algorithm in these aspects, and summarizes the results in Table 1. The table lists the key data of the algorithm in detail, such as processing speed and memory occupation, which provides an important basis for comprehensively evaluating the performance of the algorithm.

Table 1. Comprehensive Evaluation of Algorithm Performance

Algorithm Name	Accuracy (%)	Processing Speed (Frames Per Second, FPS)	Memory Footprint (MB)
Proposed Algorithm	87.5	250	150
YOLO Algorithm	85.0	300	120

Description:

Processing speed (FPS): It reflects the computational efficiency of the algorithm. YOLO algorithm processes 250 frames per second, while the proposed algorithm processes 300 frames per second, which shows that the proposed algorithm has a slight advantage in speed.

Memory occupation (MB): The memory space occupied by the algorithm in the running process is recorded. YOLO algorithm occupies 150MB and the proposed algorithm occupies 120MB, indicating that the proposed algorithm is more efficient in memory occupation.

4. Conclusions

This paper focuses on complex scene understanding and object detection, and puts forward a series of innovative algorithms and technologies. In multi-scale object detection, by introducing FPN and attention mechanism, the detection ability of the algorithm for objects of different sizes is significantly improved. For occlusion and dense scenes, this study effectively improves the detection performance by using context information and optimized processing strategy. At the same time, in terms of real-time performance and efficiency optimization, faster detection speed is realized through model pruning and hardware acceleration. Experimental results show that the proposed algorithm has achieved excellent performance on public data sets, which verifies its effectiveness and practicability.

In the future, this study will continue to explore more efficient algorithms for complex scene understanding and object detection, especially for extreme situations and complex backgrounds. At the same time, cross-modal fusion technology will be deeply studied to realize effective fusion of more types of information. In addition, we will also pay attention to the landing problems of the algorithm in practical applications, such as data acquisition, labeling and model deployment.

References

[1] Sakaridis, Christos, Dai, et al. Semantic Foggy Scene Understanding with Synthetic Data[J]. INTERNATIONAL JOURNAL OF COMPUTER VISION, 2018, 126(9):973-992.

[2] Jayachitra J, Devi K S, Satti M S K. Terahertz video-based hidden object detection using YOLOv5m and mutation-enabled salp swarm algorithm for enhanced accuracy and faster recognition[J]. Journal of supercomputing, 2024, 80(6):8357-8382.

[3] Mahalingam T, Subramoniam M. Optimal object detection and tracking in occluded video using DNN and gravitational search algorithm[J]. Soft Computing, 2020, 24(24):18301-18320.

[4] Zhou W, Wang X, Fan Y, et al. KDSMALL: A lightweight small object detection algorithm based on knowledge distillation[J]. Computer Communications, 2024, 219:271-281.

[5] Changdong W, Rui L. Object detection algorithm for indoor switchgear components in substations based on improved YOLOv5s[J]. Insight-Non-Destructive Testing and Condition Monitoring, 2024, 66(4):226-231.

[6] Wang Y, Liu X, Guo R. An object detection algorithm based on the feature pyramid network and single shot multibox detector[J]. Cluster Computing, 2022, 25(5):3313-3324.

[7] Lee J N, Cho H C. Automated Polyp Detection System in Colonoscopy using Object Detection Algorithm based on Deep Learning[J]. Transactions of the Korean Institute of Electrical Engineers, 2021, 70(1):152-157.

[8] Huang Z, Yin Z, Ma Y, et al. Mobile phone component object detection algorithm based on improved SSD[J]. Procedia Computer Science, 2021, 183(2):107-114.