ISSN 2616-5775 Vol. 7, Issue 1: 59-64, DOI: 10.25236/AJCIS.2024.070109

Application of Computer Vision Algorithms in Image Recognition and Object Detection

Herui Wang*

Faculty of Engineering, University of Waterloo, Waterloo N2L3G1, Canada
**Corresponding author*

Abstract: *Computer vision algorithms have important applications in the fields of image recognition and object detection. With the development of deep learning technology, computer vision algorithms have made significant progress in tasks such as object detection, classification, and positioning. In this study, convolutional neural networks and large-scale data sets are used for training to explore the application of computer vision algorithms in image recognition and object detection. The performance of the algorithm in target recognition and detection tasks is evaluated through feature extraction and model training of image data. The experimental results show that the accuracy rate of this algorithm is between 89% and 97%, and the computer vision algorithm has high accuracy and robustness in image recognition tasks. Through the effective training of deep learning models, the algorithm can automatically identify and classify different objects and scenes in the image.*

Keywords: *image recognition, object detection, computer vision*

1. Introduction

The application of computer vision algorithms in the field of image recognition and object detection has become the focus of people's attention. With the large-scale generation and storage of digital images, as well as the rapid development of deep learning technology, computer vision algorithms can automatically recognize and understand objects and scenes in images, which has brought important application value to many fields.

The purpose of this paper is to study and explore the application of computer vision algorithms in image recognition and object detection. By using deep learning models and large-scale data sets for training, we will evaluate the performance of the algorithm in tasks such as target detection, classification, and positioning. By optimizing and improving the algorithm, we hope to improve the accuracy, robustness and efficiency of the algorithm, so as to provide better solutions for practical applications. This article will first review the development history of computer vision algorithms and related research results, and introduce the current research status and existing challenges. Next, we will describe the research methods in detail, including the selection and preprocessing of data sets, the construction and training of deep learning models, etc. Then, we will show the experimental results and analyze and discuss them. Finally, we will summarize the main findings of the research and discuss future development directions and possible application scenarios.

2. Related work

Many scholars have conducted research on image recognition algorithms and object detection. Among them, Wang Hongyao proposed a wire rope damage recognition and detection method based on deep learning, using the target detection algorithm YOLOv5 and improving it.Fast adaptive weighted median filtering is used for image preprocessing to improve the image recognition accuracy of wire rope damage. After the improvement, the operating speed is increased to 187 ms/sheet, and the enhancement effect is good [1].Wang Keping proposed a remote sensing image recognition algorithm based on pseudo-global Swin Transformer. He construct a pseudo-global Swin transformer module to aggregate the local information of the rasterized remote sensing image into one eigenvalue, replace the pixel-based global information, and obtain global features at the cost of a smaller amount of calculation, effectively improving the model's ability to perceive all targets [2].Based on the SAR-BagNet model, Li Peng added spatial attention and coordinate attention mechanisms to the model framework, and conducted experiments on the MSTAR measured data set. The experimental results show that the

ISSN 2616-5775 Vol. 7, Issue 1: 59-64, DOI: 10.25236/AJCIS.2024.070109

spatial attention and coordinate attention mechanisms enhance the global information acquisition ability of the SAR-BagNet model, and can effectively improve the recognition accuracy and decision rationality of the model without reducing its interpretability [3].Zhou Jinwei conducted a detailed investigation on the important applications of YOLO series algorithms and compared them with existing object detection algorithms. On this basis, he summarized the characteristics of YOLO and pointed out the future development trend of YOLO algorithm [4].Hou Yueqian proposed a multi-scale object detection network based on Transformer, which uses a cross-scale embedding layer to initially embed image features; multi-branch hollow convolution is used to down sample the input, and the expansion rate of parallel branches is adjusted to make the structure have a variety of sensory fields; then, the output embedding result is processed by the residual self-attention module to construct a connection for the local and global information of the feature map, so that the attention calculation is integrated into effective multi-scale semantic information, and finally multi-scale object detection is realized [5].The accuracy rate of the above methods in image recognition and object detection has not achieved the desired effect. This paper will conduct a deeper study of image recognition and object detection through computer vision technology.

3. Method

3.1 Image recognition algorithm

In image recognition algorithms, feature extraction is a key step, which describes the image by extracting important information from the image that helps in the recognition task. The feature extraction method used in this paper is a feature representation based on deep learning. It is an advanced feature learned from the original pixel data through convolutional neural networks. These features can capture key information such as image structure, texture, and color [6-7].The classifier model is used in image recognition to map extracted features to pre-defined categories or labels. The K nearest neighbor classifier model is used to establish the mapping relationship between features and categories through the learning and training process. Because of its good classification ability, the convolutional neural network model has become the mainstream classifier model in image recognition due to its learning ability and excellent generalization performance of complex features. Feature extraction and classifier models work closely together in image recognition algorithms to achieve accurate image classification and recognition tasks by extracting meaningful features of the image and inputting them into the appropriate classifier model [8-9].This combination of methods can effectively improve the performance and accuracy of image recognition.

3.2 Object detection

Canny edge detection is used to locate the target, and Gaussian filtering is performed on the input image to smooth the image and reduce the impact of noise. Gaussian filtering is achieved by applying a Gaussian kernel to the image. Next, calculate the gradient amplitude and direction of the smoothed image, and achieve it by applying the Sobel operator (or other gradient operator) to convolve the image in the horizontal and vertical directions [10-11].In this step, the gradient amplitude image is scanned, leaving only the local maximum pixels to refine the edges, which can suppress the response of non-edge areas and make the edges thinner. The edge pixels are divided into strong edge, weak edge, and non-edge pixels by setting two thresholds. A high threshold value is used to mark strong edge pixels, a low threshold value is used to mark weak edge pixels, and non-edge pixels are excluded. Finally, by connecting the strong edge pixels and the adjacent weak edge pixels, a complete edge segment is formed. The connectivity analysis algorithm is applied to achieve it.

In the area generation method, a sliding window is used to achieve it. The size and aspect ratio of the sliding window are defined as 5*5 and 1:1, and the window size corresponds to the target size. Slide the defined window on the image in a fixed step size. The step size is set to 10. The step size determines the degree of overlap between the windows. Usually a part of the window size is selected, and the sliding window slides in all positions of the image, including horizontal and vertical directions. For each sliding window position, the features extracted by the convolutional neural network are represented from the window, and for the window determined by the classifier to contain a target, it is used as a candidate area for further processing, such as target positioning, target recognition, etc. [12-13].

ISSN 2616-5775 Vol. 7, Issue 1: 59-64, DOI: 10.25236/AJCIS.2024.070109

3.3 Algorithm optimization and improvement

MobileNet, a lightweight network structure, is used to reduce the amount of parameters and calculations of the model, and through pruning and thinning techniques, redundant parameters in the model are reduced to improve the computational efficiency of the model. For optimization, acceleration techniques such as model quantification, model distillation, and model decomposition are used to speed up the reasoning speed of the model, and different data enhancement techniques such as rotation, translation, scaling, and flipping are applied to generate diverse training samples to increase the generalization ability of the model. The algorithm uses weakly labeled and unlabeled data for training to minimize the workload of manual labeling and improve the efficiency of data utilization. The model uses a single-stage detector YOLO, which directly predicts the location and category of the target through dense anchor box/a priori box generation.Target detection is carried out through multi-scale feature maps to capture targets of different sizes. In this article, a feature pyramid is constructed to integrate features of different scales to improve the accuracy of target detection [14-15].Through the improvement of the network structure, the design of the loss function, and the attention mechanism, the performance of the first-phase detector has been further improved.

4. Results and discussion

4.1 Experimental design

This article design experiments to verify the performance and feasibility of computer vision algorithms in image recognition and object detection tasks, evaluate the performance of selected computer vision algorithms in image recognition and object detection tasks, compare the performance differences of different algorithms or parameter configurations, and verify the feasibility and applicability of algorithms in practical applications. The experimental group used the algorithm of this paper for experiments, and the control group used traditional algorithms for image recognition and object detection. Before experiments select an appropriate data set that contains a large number of image samples and corresponding labeling information for image recognition and object detection tasks, and preprocess the data set, including image size adjustment, data enhancement, etc., to ensure data consistency and availability. Use the training set to train the selected algorithm, adjust the network structure and parameter configuration, and monitor the performance indicators during the training process. The accuracy rate and the average intensive reading mean are used as the experimental evaluation indicators of this paper. Multiple experiments were carried out in the test to calculate the average performance index to reduce the random error of the experimental results.

4.2 Accuracy

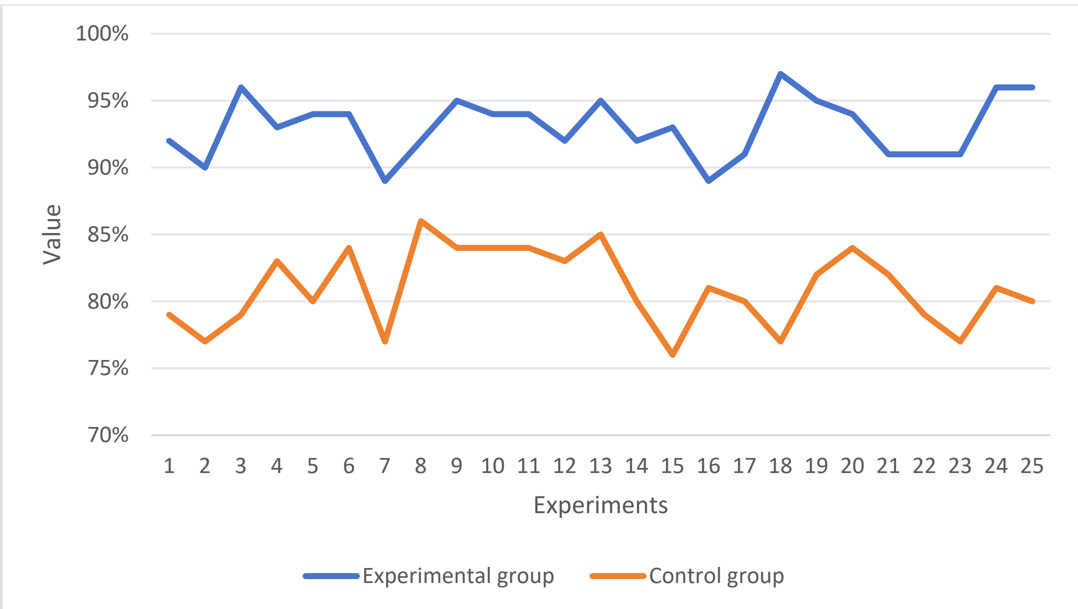

Figure 1. Accuracy rate

ISSN 2616-5775 Vol. 7, Issue 1: 59-64, DOI: 10.25236/AJCIS.2024.070109

In image recognition tasks, the accuracy rate reflects the accuracy rate of the algorithm when classifying images. It can tell us how much the model has the ability to correctly divide images into various categories. The higher accuracy rate means that the algorithm has good classification ability and can accurately identify the content in the image. In object detection tasks, the accuracy rate represents the accuracy of the algorithm when locating and identifying the target object. It considers the location and category of the target, and evaluates the performance of the algorithm by calculating the degree of overlap between the detected target and the real target. The higher accuracy rate means that the algorithm can accurately locate and identify the target object. Figure 1 is the accuracy test result.

In the experimental results of the accuracy rate, the accuracy rate of the experimental group was between 89% and 97%, the highest accuracy rate of the control group was only 87%, and the highest accuracy rate of the control group did not reach the lowest value of the experimental group, indicating that the accuracy rate of this method is much higher than that of traditional methods. This is because computer vision algorithms are usually based on deep learning models and have strong automatic feature learning capabilities. Compared with traditional feature extraction methods that require manual design, deep learning models can automatically learn more discriminating feature representations from large amounts of data, thereby improving accuracy.

4.3 Average accuracy mean

Mean Average Precision (mAP) is a commonly used performance evaluation index, which is widely used in image recognition and object detection tasks of computer vision algorithms. Map comprehensively considers the accuracy and recall rate of target detection algorithms in different categories, and is used to evaluate the overall performance of the algorithm. Figure 2 shows the experimental results of MPa:

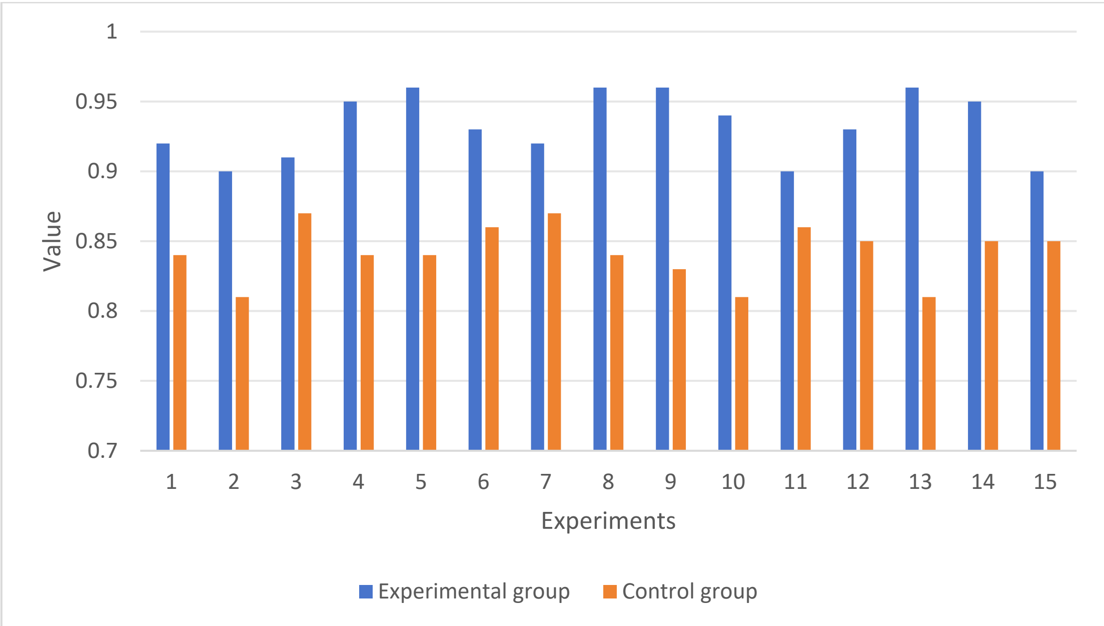

Figure 2. Average accuracy mean

In the MPa test, the MPa value of the experimental group was between 0.9-0.96, while the MPa of the control group was between 0.81-0.87, which was much smaller than that of the experimental group in this paper. The higher mAP means that the algorithm has a higher average accuracy in multiple categories and can more accurately detect and locate target objects. Compared with traditional methods, computer vision algorithms usually have better performance in target detection tasks, can identify more targets, and reduce the false detection rate. Table 1 shows the actual effect of the algorithm in this paper and the traditional algorithm when detecting objects.

ISSN 2616-5775 Vol. 7, Issue 1: 59-64, DOI: 10.25236/AJCIS.2024.070109

Table 1. Actual recognition effect

Category	Experimental group	control group
Car	0.85	0.81
Pedestrian	0.88	0.78
Bicycle	0.92	0.76
Cell phone	0.91	0.80
Fire hydrant	0.89	0.81
mAP	0.85	0.80

5. Conclusion

Computer vision algorithms are widely used in the fields of image recognition and object detection. Through the training of deep learning models and large-scale data, computer vision algorithms have made significant progress in tasks such as target detection, classification, and positioning. In terms of image recognition, computer vision algorithms can automatically recognize and classify different objects and scenes in images. Through the use of deep learning models such as convolutional neural networks (CNN), algorithms can learn from images to more discriminating feature representations, improving the accuracy and robustness of recognition In terms of object detection, computer vision algorithms can detect and locate target objects in images. Through techniques such as area proposal and anchor frame, the algorithm can filter and precisely locate areas in the image that may contain targets. The performance of computer vision algorithms in target detection tasks continues to improve, and they can accurately detect and track targets in complex scenes. The application of computer vision algorithms in image recognition and object detection is of great significance to many fields. In the fields of autonomous driving, video surveillance, industrial quality inspection, and medical imaging, computer vision algorithms can provide efficient and accurate image analysis and understanding, which has brought great convenience and benefits to people's lives and work.

References

[1] Wang Hongyao, Han Shuang, Li Qinyi. Research on the experimental method of improving the image recognition of wire rope damage of YOLOv5 [J].Computer Engineering and Applications, 2023, 59(17):99-106.

[2] Wang Keping, Zuo Xinhao, Yang Yi, Fei Shumin. Remote sensing image recognition algorithm based on pseudo-global Swin Transformer [J]. Pattern recognition and Artificial Intelligence, 2023, 36(9):818-831.

[3] Li Peng, Feng Cunqian, Hu Xiaowei. An improved interpretable SAR image recognition network [J].Journal of Air Force Engineering University, 2023, 24(4):49-55.

[4] Zhou Jinwei, Wang Jianping. Review of research on YOLO object detection algorithm [J].Journal of Changzhou Institute of Technology, 2023, 36(1): 18-23+88.

[5] Hou Yueqian, Zhang Lihong. Multi-scale object detection based on Transformer [J].Journal of Testing Technology, 2023, 37(4):342-347.

[6] Ngugi L C, Abelwahab M, Abo-Zahhad M. Recent advances in image processing techniques for automated leaf pest and disease recognition–A review[J]. Information processing in agriculture, 2021, 8(1): 27-51.

[7] Hong D, Gao L, Yao J, et al. Graph convolutional networks for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 59(7): 5966-5978.

[8] Zhu Y, Zhuang F, Wang J, et al. Deep subdomain adaptation network for image classification[J]. IEEE transactions on neural networks and learning systems, 2020, 32(4): 1713-1722.

[9] Masana M, Liu X, Twardowski B, et al. Class-incremental learning: survey and performance evaluation on image classification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(5): 5513-5533.

[10] Monga V, Li Y, Eldar Y C. Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing[J]. IEEE Signal Processing Magazine, 2021, 38(2): 18-44.

[11] Xu M, Li C, Zhang S, et al. State-of-the-art in 360 video/image processing: Perception, assessment and compression[J]. IEEE Journal of Selected Topics in Signal Processing, 2020, 14(1): 5-26.

[12] Sun Y, Xue B, Zhang M, et al. Automatically designing CNN architectures using the genetic algorithm for image classification[J]. IEEE transactions on cybernetics, 2020, 50(9): 3840-3854.

ISSN 2616-5775 Vol. 7, Issue 1: 59-64, DOI: 10.25236/AJCIS.2024.070109

[13] Hong D, Wu X, Ghamisi P, et al. Invariant attribute profiles: A spatial-frequency joint feature extractor for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(6): 3791-3808.
[14] Kuznetsova A, Rom H, Alldrin N, et al. The open images dataset v4: Unified image classification, object detection, and visual relationship detection at scale[J]. International Journal of Computer Vision, 2020, 128(7): 1956-1981.
[15] Ramesh S, Vydeki D. Recognition and classification of paddy leaf diseases using Optimized Deep Neural network with Jaya algorithm[J]. Information processing in agriculture, 2020, 7(2): 249-260.