Proceedings of the 5th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/120/2025.18913

Performance Analysis and Enhancement of Multi-Object Image Recognition Algorithms Based on Machine Learning

Pengfei Pan[1,a,*]

[1]*University of Washington, 1410 NE Campus Pkwy, Seattle, WA, 98195, US*
a. pennypan210@gmail.com
**corresponding author*

Abstract: Multi-object image recognition is a pivotal research area in computer vision with widespread applications in security monitoring, autonomous driving, medical imaging, and beyond. Recent advancements in machine learning have significantly propelled the progress of image recognition algorithms. However, challenges such as limited precision, difficulty in detecting small objects, and computational inefficiencies persist. This paper delves into a performance analysis of existing multi-object image recognition algorithms, examining their applications across diverse scenarios. By optimizing network structures, employing data augmentation, and incorporating transfer learning techniques, we propose enhanced methods to elevate algorithm performance. Experimental results demonstrate substantial improvements in accuracy, recall, and operational efficiency, underscoring the effectiveness and practicality of the proposed strategies. This study not only offers theoretical insights into the field of multi-object image recognition but also provides actionable guidance for algorithm selection and optimization in real-world applications.

Keywords: multi-object image recognition, machine learning, performance analysis, algorithm enhancement, deep learning.

1. Introduction

The role of multi-object image recognition in automation and intelligent systems is indispensable. Yet, as data scales grow and scenarios become increasingly complex, the field grapples with bottlenecks such as the detection of small objects, occlusion management, real-time performance, and robustness. In areas like security surveillance and autonomous driving, achieving both high accuracy and efficiency is imperative—demands that traditional algorithms struggle to meet. Leveraging machine learning to optimize existing algorithms for greater adaptability across scenarios has emerged as a vital research direction. Identifying technical shortcomings and proposing improvement strategies will drive breakthroughs and practical applications in this domain.

2. Performance Analysis of Multi-Object Image Recognition Algorithms

2.1. Core Technologies in Multi-Object Image Recognition

The crux of multi-object image recognition lies in accurately detecting and classifying multiple targets within complex scenes. Deep learning, particularly the hierarchical feature extraction

DOI: 10.54254/2755-2721/120/2025.18913

capabilities of convolutional neural networks (CNNs), has revolutionized this process. CNN-based region proposal networks (RPNs) have significantly bolstered detection efficiency and accuracy by generating candidate regions. Algorithms such as YOLO, Faster R-CNN, and SSD are widely utilized. YOLO excels in speed-sensitive scenarios with its end-to-end training capabilities, enabling real-time detection. Faster R-CNN, on the other hand, shines in accuracy, particularly in detecting small objects, though it lags in real-time performance. SSD achieves a balance between speed and accuracy, making it suitable for resource-constrained environments. However, these algorithms still face limitations in small object detection and adapting to complex backgrounds, leaving room for further optimization[1].

2.2. Performance Metrics for Algorithm Evaluation

Evaluating multi-object image recognition algorithms necessitates a multi-dimensional approach encompassing accuracy, efficiency, and robustness. Accuracy, as the cornerstone, is often measured using precision and recall. While precision gauges the correctness of detections, recall assesses the completeness of target identification. The F1 score offers a balance between these metrics, while mean average precision (mAP) evaluates performance across multiple categories. Efficiency metrics, including runtime, memory consumption, and hardware dependency, are equally critical, particularly in real-time applications requiring rapid response. Robustness, which measures an algorithm's stability under varying lighting, occlusions, and complex backgrounds, directly influences its practical utility. These comprehensive performance metrics provide essential benchmarks for selecting and optimizing algorithms for specific applications.

2.3. Comparative Analysis of Current Algorithms

A comparative evaluation on benchmark datasets such as COCO and PASCAL VOC reveals distinct strengths and weaknesses among mainstream algorithms. Faster R-CNN demonstrates exceptional performance in complex scenarios and small-object detection, achieving over 80% mAP, but its slower runtime limits its use to static data and high-resource environments. YOLO, with its single-stage detection design, delivers superior real-time performance, exceeding 60 frames per second—ideal for high-speed scenarios like autonomous driving and surveillance. However, YOLO's precision declines in small-object and complex-background scenarios. SSD, leveraging multi-scale feature maps, strikes a balance between detection speed and accuracy, performing well on resource-constrained mobile devices. Yet, its stability falters under extreme lighting and dynamic conditions[2].

The trade-offs between precision and efficiency underscore the diversity of requirements across application scenarios. For instance, while Faster R-CNN prioritizes precision, YOLO emphasizes speed. These distinctions highlight the need for algorithm optimization tailored to specific demands. A comprehensive analysis of these strengths and limitations offers invaluable insights for advancing the field.

3. Strategies for Enhancing Multi-Object Image Recognition Algorithms

3.1. Optimizing Network Structures

The key to improving multi-object image recognition lies in refining network structures. While existing deep learning models perform well in specific contexts, they often struggle with small objects, occlusion, and complex backgrounds. Incorporating attention mechanisms enhances the model's ability to capture salient features by weighting critical information and suppressing background noise. Lightweight network architectures, such as MobileNet and ShuffleNet, further

DOI: 10.54254/2755-2721/120/2025.18913

balance accuracy and computational efficiency, enabling deployment in mobile and real-time scenarios. Customizing the number and arrangement of feature extraction layers based on task-specific requirements enhances the model's capability to discern fine details, providing adaptable solutions for diverse applications.

3.2. Data Augmentation and Preprocessing

High-quality data is paramount for training and optimizing models. Data augmentation techniques generate diverse training samples, enhancing the model's generalization capabilities. Methods such as random cropping, rotation, and color transformation simulate real-world variations, enabling the model to perform robustly under conditions of varying lighting and perspective. Additionally, synthetic datasets created using generative adversarial networks (GANs) address data scarcity, reducing reliance on manual annotations. Preprocessing steps, including image normalization and data distribution balancing, further improve training outcomes[3].

3.3. Transfer Learning and Domain Adaptation

Transfer learning proves invaluable in small-sample scenarios. By pretraining models on large-scale datasets and transferring their parameters to target tasks, initial performance and training efficiency are significantly enhanced. This approach reduces dependency on extensive labeled data while leveraging the feature extraction capabilities of pretrained models. Domain adaptation techniques dynamically adjust model parameters to account for distributional discrepancies across scenarios, ensuring consistent performance in cross-environment applications.

3.4. Multi-Task Learning and Joint Optimization

Multi-task learning integrates detection and classification tasks within a unified framework, optimizing shared features and boosting both accuracy and efficiency. This is achieved by sharing a common feature extraction network while designing task-specific output branches. Weighted loss functions balance performance across tasks, ensuring a harmonious trade-off. Joint optimization extends to multimodal data integration, combining information from images, text, or audio to provide more comprehensive support for image recognition.

4. Experimental Validation and Results

4.1. Experimental Setup

To validate the enhanced algorithms, experiments utilized standard datasets such as COCO and PASCAL VOC, supplemented by 1,200 traffic scene images comprising four vehicle types (trucks, buses, vans, and cars). Hardware configurations included an Intel i7-6800 CPU and NVIDIA GTX 1080 Ti GPU, running on Ubuntu 16.04 with the Darknet framework. Data augmentation techniques like random cropping, rotation, and saturation adjustments enriched training samples, enhancing adaptability[4]. Initial learning rates were set at 0.001, with performance evaluations conducted after 60,000 and 70,000 iterations.

4.2. Comparative Results

The performance of original YOLOv2, YOLOv2 VOC, YOLOv3, and the enhanced YOLOv2 VOC_mul model was compared. The enhanced model achieved a precision of 99.20%, a recall rate of 95.39%, and an mAP of 88.24%. Compared to its predecessors, the enhanced model excelled in small-object detection and complex backgrounds, significantly reducing missed detections. Training

DOI: 10.54254/2755-2721/120/2025.18913

showed rapid and stable convergence, with loss values approaching zero. After 70,000 iterations, false detections were further minimized, ensuring more reliable results.

4.3. Results Analysis and Discussion

In real-world road scene tests, the enhanced model demonstrated superior adaptability, excelling in detecting small objects within complex backgrounds while effectively addressing overlapping and multi-target interference. However, average accuracy for inconspicuous targets like vans remained at 89.44%, indicating room for further improvement. The results highlight that network optimization and parameter tuning can achieve a balanced trade-off between accuracy, stability, and efficiency, providing robust support for target detection in complex scenarios[5].

5. Conclusion

Enhancing multi-object image recognition algorithms is not only a technological imperative but also a pressing demand in real-world applications seeking intelligent and efficient solutions. By refining the YOLOv2 model, this study achieved significant breakthroughs in accuracy, efficiency, and stability, offering new pathways for small-object detection and multi-object recognition under complex conditions. These empirical advancements underscore the potential of deep learning in image recognition and lay a solid foundation for broader applications. Future optimization of multi-object image recognition should prioritize data diversity, model lightweighting, and domain adaptability to meet the challenges of increasingly intricate environments and unlock new possibilities in intelligent vision systems.

References

[1] *Pal S K, Pramanik A, Maiti J, et al. Deep learning in multi-object detection and tracking: state of the art[J]. Applied Intelligence, 2021, 51: 6400-6429.*

[2] *Ahn H, Cho H J. Research of multi-object detection and tracking using machine learning based on knowledge for video surveillance system[J]. Personal and Ubiquitous Computing, 2022, 26(2): 385-394.*

[3] *Ahmed M W, Alshahrani A, Almjally A, et al. Remote Sensing Image Interpretation: Deep Belief Networks for Multi-Object Analysis[J]. Ieee Access, 2024.*

[4] *Kalake L, Wan W, Hou L. Analysis based on recent deep learning approaches applied in real-time multi-object tracking: a review[J]. IEEE Access, 2021, 9: 32650-32671.*

[5] *Kuppusamy P, Hung C L. Enriching the multi-object detection using convolutional neural network in macro-image[C]//2021 International Conference on Computer Communication and Informatics (ICCCI). IEEE, 2021: 1-5.*