Object Detection and Instance Segmentation with YOLOV8: Progress and Limitations

Law Jack Lee, Hazry Desa, Muhammad Azizi Azizan
Centre of Excellence for Unmanned Aerial Systems (COE-UAS), Universiti Malaysia Perlis, Block E, Pusat Perniagaan Pengkalan Jaya, Jalan Kangar – Alor Setar, 01000 Kangar, Perlis.

Abadal-Salam T. Hussain
Department of Medical Equipment Technology Engineering, College of Engineering Technology, Al-Kitab University, Kirkuk, Iraq.

Muhammad Hassan Tanveer
Department of Robotics and Mechatronics Engineering, Kennesaw State University, Marietta, GA, 30067, USA.

Email: hazry@unimap.edu.my
www.unimap.edu.my

Abstract

This research employs object detection and instance segmentation algorithms to distinguish between objects and backgrounds and to interpret the detected objects. The YOLOV8 (You Only Look Once) framework and COCO dataset are utilized for detecting and interpreting the objects. Additionally, the accuracy of detection, segmentation, and interpretation is tested by placing objects at various distances from the camera. The algorithm's performance was evaluated, and the results were documented. In the experiments, a sample of 11 objects was tested, and 8 of them were successfully detected at distances of 45cm, 75cm, 105cm, and 135cm. For instance, segmentation, segmentation maps appeared clean when detecting a single object but faced challenges when multiple objects overlapped.

Keywords: Image Processing YoloV8, COCO, Segmentation, Object detection.

1. Introduction

Traditional methods in object detection rely on hand-crafted features and rules to identify and categorize objects, whereas machine learning-based approaches involve algorithms trained on extensive labeled image datasets. Notably, machine learning-based methods offer the advantage of handling complex and diverse environments by learning and adapting to new data [1]. However, they can be computationally demanding and require substantial labeled data for training. In the past few years, there have been notable strides in object detection algorithms, especially with the rise of deep learning-centered methods that have demonstrated cutting-edge capabilities on a variety of benchmarks. These methods utilize neural networks to learn features and classify objects, successfully finding applications in various domains [2], [3]. Nevertheless, prevailing monitoring systems might exhibit limited adaptability in computer vision, leading to difficulties in accurately detecting and classifying objects or features within visually intricate settings, including those characterized by low illumination, high contrast, or complex backgrounds [4]. Furthermore, current monitoring setups often heavily rely on human operators to analyze images and spot potential issues [5]. This manual procedure can be time-intensive and prone to errors, as human operators might encounter challenges in consistently and precisely interpreting images or detecting subtle anomalies [4]. To confront these challenges, this project introduces an object detection and self-interpreting application designed to augment monitoring systems.

2. Related Work

The research landscape in computer vision and deep learning has witnessed significant advancements in recent years, particularly in the domain of object detection and segmentation. Notably, Kousik et al. proposed a hybrid Convolution Recurrent Neural Network for improved salient object detection [9]. Additionally, Mekala et al. introduced a Deep Reinforcement Learning (DRL) based 4-r Computation Model for Object Detection on Roadside Units (RSU) using LiDAR on the Internet of Things (IoT) context [10]. Deep learning and image processing techniques for cancer diagnosis were effectively employed by Prassanna et al. [11], while Ghantasala et al. focused on texture recognition and image smoothing for microcalcification and mass detection in abnormal regions [12]. In the context of the Internet of Medical Things, Manimurugan et al. utilized a deep belief neural network for effective attack detection in smart environments [13]. Furthermore, Natarajan et al. applied fully convolutional deep neural architecture for the segmentation of nuclei in histopathology images [14]. Lastly, Selvaraj et al. proposed an optimal virtual machine selection approach using swarm intelligence for anomaly detection [15]. These diverse works collectively contribute to the broader understanding and application of advanced techniques in the realm of computer vision and deep learning.

3. Methodology

Overall Object Detection and Instance Segmentation Segmentation is a specific type of image analysis that focuses on identifying individual instances of objects and outlining their boundaries [6]. Two essential computer vision tasks, semantic segmentation, and object detection are closely related to instance segmentation [7]. Semantic segmentation seeks to correctly assign the appropriate object category to every individual pixel present within an image. However, it fails to differentiate among separate instances of identical object classes. For instance, if there are three dogs in an image, semantic segmentation will only identify pixels belonging to one of these three dogs.

Conversely, object detection involves predicting both the object's bounding box and its corresponding class for every instance of an object present in the image. However, it does not provide per-pixel identification for each object instance. Hence, object instance segmentation is notably more challenging than semantic segmentation and object detection because it seeks to both identify and label each object instance on a per-pixel basis [7].

To create a dataset, for example, one can collect images of cats and dogs, organizing them into two folders: "cats" and "dogs." Each folder should contain a similar number of images. In each folder, a total 80% of the images are allocated as training images, and the remaining 20% are used as test images. Annotations are required for the training images, involving manually drawing bounding boxes and segmenting the dogs or cats from the background. These annotations produce coordinates that are recorded in a blank text file, essential for the machine's ability to locate objects during image training and testing. Fig. 1 displays an example of object detection, with a bounding box around the Region of Interest (ROI), labelled with its name, and segmented with a coloured mask layer within the ROI, utilizing the COCO dataset [8].

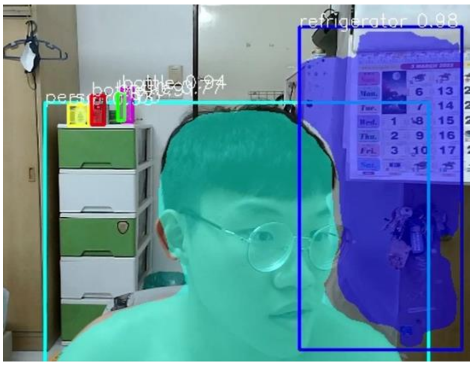
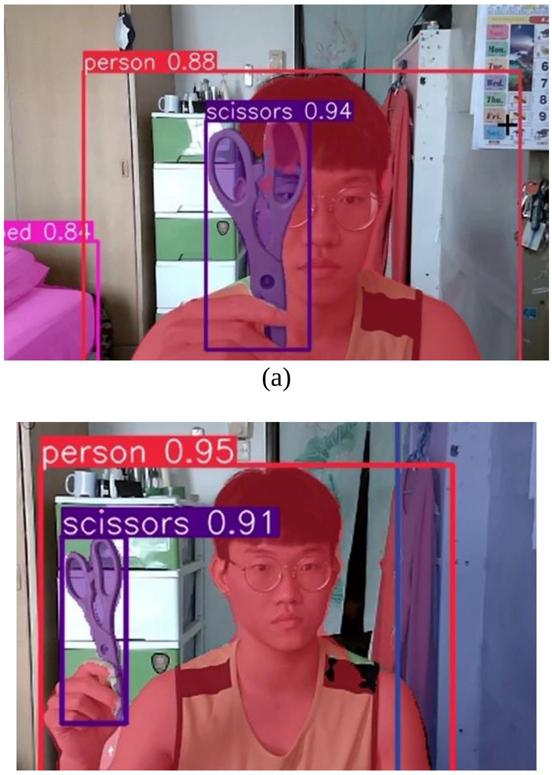


Fig. 1. Object detection and instance segmentation.

4. Results and Discussion

4.1 Object Detection and Self-Interpreting Application Within Different Distance.

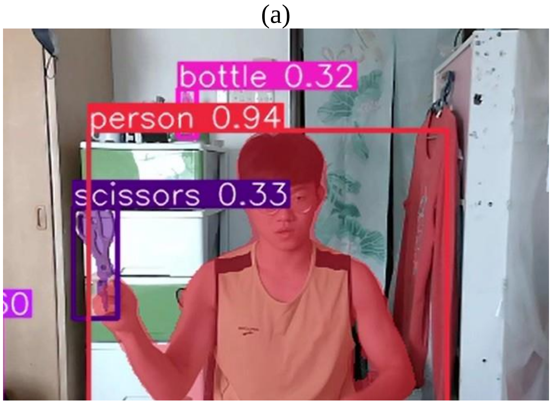
To test the ability of detection and interpreting application, the experiment was carried out in room. 11 objects were selected, and each object was placed in 45cm, 75cm, 105cm and 135cm apart from the webcam. The results were recorded in Table 1 below. A tick indicates that the detection was successful, and a cross indicates that the detection was unsuccessful.



(b)

Fig. 2. (a) Scissor at 45 cm distance and (b) scissor at 75 cm distance from the camera.

(a)



(b)

Fig. 3. (a) Scissor at 105 cm distance and (b) scissor at 135 cm distance from the camera.

Fig. 2 and Fig. 3 present examples of various distances captured by the webcam. The figures show a pair of scissors, accurately detected, and identified by the system. Most large objects, such as a person, cell phone, keyboard, and bowl, were successfully detected and identified. However, at distances of 105 cm and 135 cm, the objects like the mouse, fork, and spoon were not recognized. This lack of recognition was primarily due to object occlusion. When the fork and spoon were held at 105 cm, the student's fingers obstructed part of the objects, making it challenging for the Yolo model to detect them. Furthermore, the webcam's resolution also impacted the Yolo model's detection rate. Although the webcam supported 1080p Full HD video quality, the output was compressed to 640px x 640px for smoother video operation. It is essential to note that this experiment was conducted on a CPU-only machine, resulting in significantly lower performance compared to using a GPU-supported machine.

Instance Segmentation Within Different Distances. The results obtained from the YOLOV8 model for instance segmentation were observed and collected. When a single object was detected at distances of 45cm, 75cm, 105cm, and 135cm, the segmentation maps of a person appeared to be clean, as evident from Fig. 2 and Fig. 3. However, in cases where multiple objects overlapped, the model encountered difficulties in accurate segmentation. Fig. 4 and Fig. 5 demonstrate the inconsistent segmentation of a keyboard at different distances. While the keyboard could be segmented successfully at 45cm when accompanied by a human, it failed to do so at 75cm, 195cm, and 135cm distances. Such limitations arise from the detection and segmentation networks, as overlapping objects pose challenges in separation and precise localization. The output image may look acceptable, but it merely represents the network's best estimation of object locations. When multiple objects overlap, the model may struggle, resulting in inaccurate bounding boxes or masks, which can sometimes lead to incorrect labels.

(a)

(b)

Fig. 4. (a) 45 cm distance (b) 75 cm distance away from the camera.

(a)

(b)

Fig. 5. (a) 105 cm distance (b) 135 cm distance away from the camera.

Table 1. Instances of successful and unsuccessful attempts of being detected.

Objects	Distance away from camera (cm) on UAV				
	45	75	105	135	Percentage (%)
Scissors	✓	✓	✓	✓	100
Person	✓	✓	✓	✓	100
Bottle	✓	✓	✓	✓	100
Cup	✓	✓	✓	✓	100
Cellphone	✓	✓	✓	✓	100
Mouse	✓	✓	×	×	50
Keyboard	✓	✓	✓	✓	100
Bowl	✓	✓	✓	✓	100
Spoon	✓	✓	×	×	50
Fork	✓	✓	×	×	50
Toothbrush	✓	✓	✓	×	75

5. Conclusion

The experiment involving object detection and a self-interpreting system was carried out successfully, employing the YOLOV8 framework and the COCO dataset to collect the results. Out of the 11 objects that underwent testing, 8 of them were effectively identified at different distances: specifically, 45cm, 75cm, 105cm, and 135cm. This led to the algorithm used in the experiment achieving an accuracy rate of 72% in terms of object detection and interpretation. However, there were instances where 4 objects failed to be detected accurately at distances of 105cm and 135cm. This was attributed to the lower image quality at these distances. Notably, the segmentation maps displayed clean results when a single object was detected, but inconsistency in segmentation maps arose when multiple objects overlapped, mainly due to the limitations of the detection and segmentation network. In conclusion, while the experiment demonstrated promising results with a 72% accuracy in object detection and interpretation, challenges persisted in cases of poor image quality and overlapping objects, affecting the segmentation performance.

Acknowledgements

This work was funded by the Universiti Malaysia Perlis (UniMAP) under the Commercialization Grant 9001-00748.

References

1. H. Ishibuchi, T. Nakashima and T. Kuroda, A Fuzzy Genetics-Based Machine Learning Method for Designing Linguistic Classification Systems with High Comprehensibility, ICONIP'99. ANZIIS'99 & ANNES'99 & ACNN'99, 6th International Conference on Neural Information Processing, August, 2002.
2. J. Han, D. Zhang, G. Cheng, N. Liu and D. Xu, Advanced Deep-Learning Techniques for Salient and Category-Specific Object Detection: A Survey, IEEE Signal Processing Magazine, Vol. 35(1), January, 2018, pp. 84-100.
3. Z. -Q. Zhao, P. Zheng, S. -T. Xu and X. Wu, Object Detection with Deep Learning: A Review, IEEE Transactions on Neural Networks and Learning Systems, Vol. 30(11), November, 2019, pp. 3212-3232.
4. S. Gundu, H. Syed and J. Harikiran, Human Detection in Aerial Images using Deep Learning Techniques, 2nd International Conference on Artificial Intelligence and Signal Processing (AISP), April, 2022, pp. 1-10.
5. R. A. Suárez Fernández, J. L. Sanchez-Lopez, C. Sampedro, H. Bavle, M. Molina and P. Campoy, Natural User Interfaces for Human-Drone Multi-Modal Interaction, International Conference on Unmanned Aircraft Systems (ICUAS), July, 2016, pp. 1013-1022.
6. A. M. Hafiz and G. M. Bhat, A Survey on Instance Segmentation: State of the Art, International Journal of Multimedia Information Retrieval, Vol. 9(3), July, 2020, pp. 171-189.
7. L. Ye, Z. Liu and Y. Wang, Depth-Aware Object Instance Segmentation, IEEE International Conference on Image Processing (ICIP), February, 2018, pp. 325-329.
8. T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Doll´ar and C. L. Zitnick, Microsoft COCO: Common Objects in Context, Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September, 2014.
9. N. Kousik, Y. Natarajan, R. A. Raja, S. Kallam, R. Patan and A. H. Gandomi, Improved Salient Object Detection Using Hybrid Convolution Recurrent Neural Network, Expert Systems with Applications, Vol. 166, March, 2021.
10. M. S. Mekala, R. Patan, A. H. Gandomi, J. H. Park and H. Y. Jung, A DRL based 4-r Computation Model for Object Detection on RSU using LiDAR in IloT. IEEE Symposium Series on Computational Intelligence (SSCI), January, 2022, pp. 01-08.
11. J. Prassanna, R. Rahim, K. Bagyalakshmi, R. Manikandan and R. Patan, Effective Use of Deep Learning and Image Processing for Cancer Diagnosis, Deep learning for Cancer Diagnosis, September, 2020, pp. 147-168.
12. G. P. Ghantasala, S. Kallam, N. V. Kumari and R. Patan, Texture Recognition and Image Smoothing for Microcalcification and Mass Detection in Abnormal Region, International Conference on Computer Science, Engineering and Applications (ICCSEA), July, 2020.
13. S. Manimurugan, S. Al-Mutairi, M. M. Aborokbah, N. Chilamkurti, S. Ganesan and R. Patan, Effective Attack Detection in Internet of Medical Things Smart Environment Using a Deep Belief Neural Network, IEEE Access, Vol. 8, April, 2020, pp. 77396-77404.

14. N. A. Natarajan, M. S. Kumar, R. Patan, S. Kallam and M. Y. N. Mohamed, Segmentation of Nuclei in Histopathology Images Using Fully Convolutional Deep Neural Architecture, International Conference on Computing and Information Technology (ICCIT-1441), November, 2020, pp. 319-325.
15. A. Selvaraj, R. Patan, A. H. Gandomi, G. G. Deverajan and M. Pushparaj, Optimal Virtual Machine Selection for Anomaly Detection Using a Swarm Intelligence Approach, Applied Soft Computing Journal, Vol. 84, November, 2019.

Authors Introduction

Mr. Law Jack Lee

He completed his bachelor’s degree in Mechatronic Engineering from Universiti Malaysia Perlis and currently works as a Mechanical Design Engineer at Pentamaster Equipment Manufacturing Sdn Bhd in Penang, Malaysia.

Prof. Dr. Hazry Desa

He earned his Bachelor of Mechanical Engineering from Tokushima University, Japan, and proceeded to pursue his Ph.D. at the Artificial Life and Robotics Laboratory, Oita University, Japan, in 2003. Presently, he serves as the director at the Center of Excellence for Unmanned Aerial Systems (COE-UAS), Universiti Malaysia Perlis, Malaysia.

Dr. Muhammad Azizi Azizan

He received his PhD in Civil Engineering from the Universiti Malaysia Perlis. He is currently a senior lecturer in the same institution. He is head of Project Integration & Management (PIM) at Centre of Excellence for Unmanned Aerial System (COE-UAS), Universiti Malaysia Perlis.

Dr. Abadal-Salam T. Hussain

He is currently a Ph.D. faculty assistant professor and head of the department for Medical Instrumentation and Technique Engineering at Alkitab University. Previously, he was a staff member in the Faculty of Electrical Engineering Technology at Universiti Malaysia Perlis (UniMAP), Malaysia.

Dr. Muhammad Hassan Tanveer

He is an Assistant Professor of Robotics and Mechatronics Engineering at Kennesaw State University. Prior to that, he was a Post Doctoral Research Associate at Virginia Tech, USA where he was working on development of Bio-sonar sensor for UAVs navigations. He received his Ph.D. in Robotics and Autonomous Sytems with specilization in 'Collaboration of Heterogenous Team of Robots' from University of Genova, Italy, in 2019.