

Recently, YOLO has been adapted for instance segmentation, enabling precise boundary delineation for detected objects. Fig.1. illustrates the architecture of YOLO, beginning with a 448x448x3 input image. The image is processed through convolutional and max pooling layers, reducing the feature map to 7x7x1024. This is followed by two fully connected layers, resulting in a 7x7x30 output tensor that represents bounding boxes, objectness scores, and class probabilities. Garbage detection example shown in Fig.2.



Fig.2. Image segmentation based on garabge dataset

2.2 Track Handling and State Estimation

The Kalman Filter is currently used for handling track estimation. In this context, we assume a general tracking scenario that can be expanded as follows: the camera used for tracking is uncalibrated, and ego-motion data is unavailable. Such a situation presents significant challenges for the Kalman Filter prediction framework, yet it is a common scenario in recent object tracking benchmarks [15]. The prediction equation of the Kalman Filter is defined as follows:

$$\hat{x}_k^- = A\hat{x}_{k-1} + Bu_{k-1} \quad (1)$$

$$\hat{P}_k^- = AP_{k-1}A^T + Q \quad (2)$$

Update equations are defined as follows:

$$K_k = \frac{P_k^- C^T}{C P_k^- C^T + R} \quad (3)$$

$$\hat{x}_k = \hat{x}_k^- + K_k(y_k - C\hat{x}_k^-) \quad (4)$$

$$P_k = (I - K_k C)P_k^- \quad (5)$$

Where K_k , \hat{x}_k , P_k , I is the Kalman gain matrix, the optimum filter value, To delve deeper into our tracking algorithm, we define a four-dimensional state space consisting of (x, y, v_x, v_y) Each component is broken down as follows: (x, y) represent the position of the object in the x and y planes of the image frame, while (v_x, v_y) denote the velocity of the object along the x and y axes. The Kalman Filter we use is a standard implementation with a constant velocity motion model, and the observation model is also linear. Expanding on our observation model, (x, y) positional information is extracted from the centroid coordinates of the detected object provided by YOLO. This positional data is then fed into the filter as a measurement update.

2.3 Data Assignment Algorithms.

When dealing with a multi-object tracking problem, data association is crucial. This is because we have multiple detected object position data and multiple Kalman Filter-predicted object positions. To ensure continuous prediction and updating of newly incoming positional data, we need to match these two sets of information. To address this challenge, we introduce a cost matrix and the Hungarian algorithm, which work together to calculate matching pairs between predicted object positions and newly detected positions.

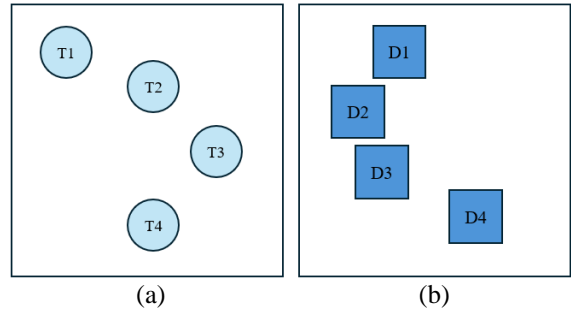


Fig.3. Diagram example of associating track and detections. (a) Tracks at time t-1, (b) Detection at time t

To determine the optimal assignment between tracks and detections, a cost matrix is crucial for representing the cost of each possible association between track positions and detections [16]. Various metrics can be used to calculate these costs; however, in our experiment, we opted to use the Euclidean distance. The Hungarian algorithm processes the cost matrix as input, performing a series of operations to simplify the assignment process [16]. Table.1. illustrates the optimal assignment, where cells with 1s indicate the mapping of tracks to detections. Fig.3. and Fig.4. demonstrates the association of objects across consecutive frames. For coding convenience, we chose to use the SciPy library, which provides the linear_sum_assignment() method. This method implements the Hungarian algorithm and returns the row and column indices corresponding to the optimal assignment [16].

| | D1 | D2 | D3 | D4 |
|----|----|----|----|----|
| T1 | 1 | 0 | 0 | 0 |
| T2 | 0 | 0 | 1 | 0 |
| T3 | 0 | 1 | 0 | 0 |
| T4 | 0 | 0 | 0 | 1 |

Table.1. Optimal assignment for the cost matrix

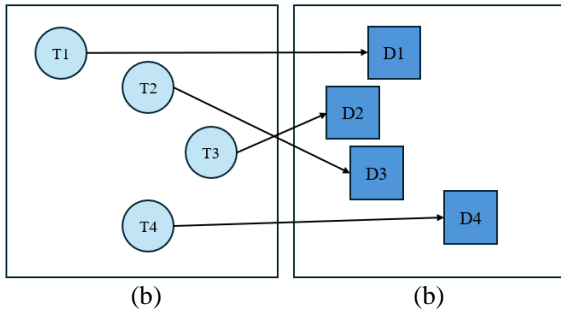


Fig.4. Final mapping example of track and detections. (a) Tracks at time t-1, (b) Detection at time t

3. Results and Discussion

3.1 Dataset

The dataset used in this study was created by combining the publicly available TACO dataset [17], which focuses on general trash, with images captured by us in Kanezaki, Munakata, Fukuoka, Japan. TACO is a comprehensive dataset containing diverse examples of trash in natural environments, featuring various locations beyond beaches, with a total of 1,500 images. Additionally, 97 images were specifically taken at cleanup sites to supplement the dataset. To improve its effectiveness in dynamic environments, we enhanced the dataset with various annotation techniques. This was particularly important given the challenges of working in uncontrolled outdoor conditions, such as varying lighting throughout the day and motion blur caused by the beach cleaning robot's movement. The final dataset consists of 3,845 images.

3.2 Object tracking and Identity switch

The primary goal of our object tracking program was to reduce identity switches caused by detection loss in the YOLOv8 detection algorithm. Detection loss is a common issue in object detection programs that rely on RGB image processing. One major cause of this problem is occlusion, along with changes in illumination and object size between frames due to different camera perspectives. Since we are working with 2D images, the information captured by the camera is inherently limited compared to 3D imaging, as it only provides a 2D perspective. This limitation introduces several points of potential error in object classification and detection. For instance, in one frame at time t-1, the algorithm might detect a PET bottle, but at time t, the image of the bottle could be distorted (e.g., sheared), causing the detection algorithm to fail to recognize it. At time step t+1, the algorithm might detect the same bottle again, but without a robust tracking program, it would treat the bottle as a new instance, leading to false object counts. This issue is referred to as an identity switch. Since the objective of

our research is to develop a program for garbage tracking and counting, addressing this issue is our top priority.

| Garbage Type | Detection loss | ID switch |
|--------------|----------------|-----------|
| PET bottle | 18 | 0 |
| Styrofoam | 21 | 1 |
| Wood | 7 | 1 |
| Shoe | 10 | 0 |

Table.2. Results of ID switch quantities compared to detection loss across multiple frames

As shown in Table 2, after implementing the tracking algorithm, the number of identity switches was significantly reduced compared to detection losses across multiple frames. The test was performed on the researchers' laboratory laptop, equipped with an Intel Core i9-12900H processor running at 2.5 GHz, 16 GB of RAM, and an NVIDIA GeForce RTX 3070 Ti Laptop GPU. The test conducted by slowing down the object detection program to 3 frames per second and observing each frame to determine how often the algorithm failed to continuously detect the same object as it moved across the frame. It was observed that the detection algorithm experienced frequent losses. Despite this, the Kalman Filter effectively used its motion model to predict the possible positions of objects accurately. Once a prediction was made, the cost matrix calculated the association cost between each track and detected object using Euclidean distance. Finally, the Hungarian algorithm was applied in the last step to match each track with a detected instance. Overall, the performance of the tracking algorithm met our requirements for the garbage tracking and counting application.

4. Conclusion

This research aims to develop a garbage tracking and counting algorithm using a Kalman Filter for tracking and predicting objects' future positions in cases of detection loss across multiple video frames. The algorithm incorporates a cost matrix and the Hungarian algorithm for data association. The YOLOv8 detection algorithm served as the foundation for the tracking program, and the detection model was trained on our custom dataset. The final results demonstrated satisfactory performance, with a small number of identity switches. For future work, we aim to implement fuzzy logic to make the maximum Euclidean distance threshold adaptive. This improvement would account for uncertainties in the detected position and the reappearance probability of each object in the multi-object tracking algorithm.

References

1. Wojke, N., Bewley, A., & Paulus, D. (2017, September). Simple online and realtime tracking with a deep association metric. In *2017 IEEE international conference on image processing (ICIP)* (pp. 3645-3649). IEEE.
2. Ciaparrone, G., Sánchez, F. L., Tabik, S., Troiano, L., Tagliaferri, R., & Herrera, F. (2020). Deep learning in video multi-object tracking: A survey. *Neurocomputing*, 381, 61-88.
3. B. Yang and R. Nevatia, "An online learned CRF model for multi-target tracking," in *CVPR*, 2012, pp. 2034-2041.
4. B. Yang and R. Nevatia, "Multi-target tracking by on line learning of non-linear motion patterns and robust appearance models," in *CVPR*, 2012, pp. 1918-1925.
5. A. Andriyenko, K. Schindler, and S. Roth, "Discrete continuous optimization for multi-target tracking," in *CVPR*, 2012, pp. 1926-1933.
6. A. Milan, K. Schindler, and S. Roth, "Detection- and trajectory-level exclusion in multiple object tracking," in *CVPR*, 2013, pp. 3682-3689.
7. L. Zhang, Y. Li, and R. Nevatia, "Global data association for multi-object tracking using network flows," in *CVPR*, 2008, pp. 1-8.
8. H. Pirsiavash, D. Ramanan, and C. C. Fowlkes, "Globally-optimal greedy algorithms for tracking a variable number of objects," in *CVPR*, 2011, pp. 1201-1208.
9. J. Berclaz, F. Fleuret, E. Turetken, and P. Fua, "Multiple object tracking using k-shortest paths optimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 9, pp. 1806-1819, 2011.
10. Kalman, R. E. (1960). A new approach to linear filtering and prediction problems.
11. Zhu, X., Rochman, C. M., Hardesty, B. D., & Wilcox, C. (2024). Plastics in the deep sea—A global estimate of the ocean floor reservoir. *Deep Sea Research Part I: Oceanographic Research Papers*, 206, 104266.
12. Ogawa, S., Jie, T. C., Tomokawa, T., Geiser, S., Chumkamon, S., Tominaga, A., & Hayashi, E. (2023). Development of Beach Litter Detection System using Deep Learning on Beach Clean-up. *Journal of Advances in Artificial Life Robotics*, 3(4), 238-241.
13. Kuhn, H. W. (1955). The Hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2), 83-97.
14. Redmon, J., & Farhadi, A. (2017). YOLO9000: better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7263-7271).
15. A. Milan, L. Leal-Taixé, I. Reid, S. Roth, and K. Schindler, "Mot16: A benchmark for multi-object tracking," arXiv preprint arXiv:1603.00831, 2016.
16. R. Sadli, *Beginner's Guide to Multi-Object Tracking with Kalman Filter*. 2023.
17. "TACO Dataset", <http://tacodataset.org/>

Authors Introduction

Mr. Rut Yatigul



He received his Bachelor's degree in Engineering in 2023 from Robotics and Automation Engineering, King Mongkut's University of Technology North Bangkok in Thailand. He is currently a Master student in Kyushu Institute of Technology at Hayashi Laboratory and conducts research about Robotics and AI.

Mr. Tan Chi Jie



He received his Bachelor of Engineering Electronics Majoring in Robotics and Automation from the Faculty of Engineering, Multimedia University, Malaysia in 2020 and received masters in 2023 from Kyushu Institute of Technology at Hayashi Laboratory. He is currently a PhD student at Kyushu Institute of Technology and conducts research at Hayashi Laboratory.

Mr. Prem Gamolped



He received bachelor degree in Engineering in 2021 from Robotics and Automation Engineering, King Mongkut's University of Technology North Bangkok in Thailand and received masters in 2022 from Kyushu Institute of Technology at Hayashi Laboratory. He is currently a PhD student in Kyushu Institute of Technology at Hayashi Laboratory and conducts research about Robotics and AI.

Prof. Eiji Hayashi



Prof. Eiji Hayashi is a professor in the Department of Intelligent and Control Systems at Kyushu Institute of Technology. He received the Ph.D. (Dr. Eng.) degree from Waseda University in 1996. His research interests include Intelligent mechanics, Mechanical systems and Perceptual information processing. He is a member of The Institute of Electrical and Electronics Engineers (IEEE) and The Japan Society of Mechanical Engineers (JSME).