IAES Journal of Electrical and Electronics Engineering
Vol. 11, No. 2, December 2024, pp. 43~50
ISSN: 2338-7432, DOI: 10.11591/jeee.v11i2.pp43-50

CCTV-based real-time detection of illegal waste dumping in residential areas using YOLOv8 and object tracking

Lina Handayani[1], Tole Sutikno[2], Watra Arsadiando[3]
[1]Faculty of Public Health, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
[2]Faculty of Industrial Technology, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
[3]Embedded Systems and Power Electronics Research Group, Yogyakarta, Indonesia

Article Info

Article history:

Received Apr 7, 2022
Revised Nov 23, 2022
Accepted Mar 24, 2023

Keywords:

CCTV surveillance
Computer vision
Illegal waste dumping
Multi-object tracking
Object detection
Urban sanitation
YOLOv8

ABSTRACT

Illegal waste dumping in residential areas remains a persistent urban sanitation challenge, contributing to environmental degradation, blocked drainage systems, unpleasant surroundings, and potential public health risks. Conventional monitoring approaches, which rely heavily on community complaints and periodic inspections, often fail to identify disposal activities promptly. Recent advances in computer vision provide opportunities to transform existing closed-circuit television (CCTV) infrastructure into intelligent environmental monitoring systems. This review examines the integration of YOLOv8-based object detection and multi-object tracking for identifying illegal waste dumping activities in real time. Particular attention is given to the relationship between waste detection, human–object interaction, temporal event recognition, and the operational requirements of urban sanitation management. The review discusses the YOLOv8 architecture, tracking algorithms, video processing pipelines, dataset development, performance evaluation, and practical deployment considerations. It also examines challenges associated with occlusion, lighting variations, camera positioning, false detections, privacy, and computational efficiency. The proposed conceptual framework emphasizes the importance of distinguishing actual dumping events from the passive presence of waste in public spaces. Ultimately, integrating intelligent CCTV surveillance with municipal waste management operations could support faster incident reporting, improved sanitation supervision, and evidence-based environmental governance.

This is an open access article under the CC BY-SA license.

1. INTRODUCTION

Rapid urbanization and population growth have increased the pressure on municipal solid waste management systems, particularly in densely populated residential areas. Inadequate waste collection services, insufficient disposal facilities, and inconsistent community participation can encourage unauthorized waste disposal in streets, vacant land, drainage channels, and other public spaces. These practices create environmental problems that extend beyond visual pollution, including soil and water contamination, unpleasant odors, and conditions that may support disease-vector proliferation [1], [2]. Consequently, illegal waste dumping should be considered both an environmental management problem and a public health concern.

Conventional approaches to identifying illegal dumping generally depend on public complaints, field inspections, and periodic observations by sanitation personnel. Although these approaches remain necessary, their effectiveness is constrained by limited monitoring coverage, delayed reporting, and the increasing number of locations requiring supervision. Closed-circuit television (CCTV) infrastructure already installed for neighborhood security offers an opportunity to improve environmental monitoring without necessarily introducing an entirely separate camera network. However, conventional CCTV monitoring still

depends substantially on human operators who must continuously interpret video streams and recognize relevant activities [3], [4].

Recent developments in deep learning have substantially improved the capabilities of automated visual recognition systems. Object detection algorithms can identify and localize multiple objects within individual video frames, while object tracking techniques establish associations between detected objects across consecutive frames [5], [6]. The YOLO family has become particularly relevant for real-time applications because it combines object localization and classification within a single detection framework. YOLOv8, introduced in 2023, represents an important development in this architecture family, providing a practical foundation for developing efficient visual monitoring applications [7].

Nevertheless, detecting waste objects is not equivalent to detecting illegal dumping activities. A camera may observe garbage that has existed for several hours or days without providing evidence of when or how it was deposited. Similarly, the presence of a person near a waste pile does not necessarily indicate an illegal disposal event. This distinction identifies an important research gap between static waste detection and temporal recognition of disposal behavior.

This review therefore examines how YOLOv8 and multi-object tracking can be integrated into a CCTV-based monitoring framework for residential environments. The discussion connects computer vision techniques with the operational requirements of urban sanitation management. Particular emphasis is placed on recognizing the sequence of events associated with waste disposal rather than simply identifying garbage in an image. The review also identifies technical, institutional, and environmental considerations that should guide the development of practical monitoring systems.

2. URBAN SANITATION PROBLEMS AND ILLEGAL WASTE DUMPING

2.1. Environmental and public health implications

Illegal waste dumping creates localized environmental pressures that can accumulate into broader urban sanitation problems. Waste deposited in residential streets may obstruct drainage channels, contribute to localized flooding, and create favorable conditions for insects and rodents. Organic waste can decompose and produce unpleasant odors, while improperly discarded plastics and other materials may remain in the environment for extended periods. The environmental consequences of unmanaged waste disposal are particularly concerning in communities with limited sanitation infrastructure and inadequate collection services [1], [2].

The relationship between illegal dumping and public health is influenced by the type of waste, disposal location, environmental conditions, and duration of exposure. Decomposing organic materials may attract disease vectors, whereas hazardous household materials can introduce chemical exposure pathways.

Waste accumulation near residential buildings, waterways, and food preparation areas can increase community exposure to environmental hazards. Therefore, timely detection and removal should be integrated into broader municipal sanitation programs rather than treated solely as a neighborhood cleanliness activity [8].

2.2. Limitations of conventional monitoring

Conventional waste monitoring generally identifies problems after waste has accumulated at a particular location. This approach provides limited information about the individuals, vehicles, or activities associated with the disposal process. Manual surveillance is also difficult to maintain continuously because municipal personnel must cover multiple locations and respond to other operational responsibilities. These limitations demonstrate the need for monitoring technologies capable of identifying suspicious activities as they occur. An intelligent CCTV system can extend conventional monitoring by analyzing video streams automatically.

Instead of requiring operators to observe every camera continuously, computer vision algorithms can identify relevant objects and generate notifications when predefined conditions are satisfied. The resulting information can include the incident location, timestamp, detected waste category, and associated video evidence. Such information can improve the coordination of sanitation personnel and support more consistent documentation of illegal dumping incidents [9], [10].

3. COMPUTER VISION FOR REAL-TIME WASTE DETECTION

3.1. Object detection approaches

Object detection is a fundamental computer vision task involving the identification and localization of objects within images or video frames. Modern approaches generally include two-stage detectors and single-stage detectors, which offer different balances between detection accuracy and computational

efficiency. Two-stage approaches, including faster R-CNN, first generate candidate object regions and subsequently classify them. Single-stage approaches, including SSD and YOLO, directly predict object locations and categories, making them attractive for applications requiring rapid inference [11]-[14].

The YOLO architecture is particularly relevant to continuous CCTV monitoring because it processes image frames through a unified detection network. Its development has progressively improved detection accuracy, feature extraction, and computational efficiency. YOLOv4 and YOLOv7 illustrate advances in training strategies and network design that contributed to the evolution of real-time object detection [15], [16]. The YOLO family has also been examined extensively in recent literature, including architectural developments up to YOLOv8 [7].

3.2. The relevance of YOLOv8

YOLOv8 is suitable as a candidate detector for residential waste monitoring because it supports object detection and other computer vision tasks within a common framework. Its anchor-free detection design and decoupled prediction head are relevant to applications involving objects with different dimensions and appearances. The model can be configured with different size variants to accommodate different computational resources and monitoring requirements. A smaller configuration may be appropriate for edge devices, whereas larger configurations can be evaluated when greater processing capacity is available [7].

The detection classes should reflect the environmental characteristics of the target residential area. Relevant categories may include plastic bags, cardboard, household garbage, furniture, construction debris, and accumulated waste piles. However, the model should distinguish between individual waste items and larger dumping accumulations because these objects have different operational meanings. A waste bag near an authorized collection point, for example, should not automatically be classified as an illegal dumping incident.

As indicated in Table 1, YOLOv8 provides a promising starting point for developing an integrated waste monitoring system. Nevertheless, architectural advantages alone cannot guarantee reliable detection in residential environments. Performance depends on image quality, training data diversity, object dimensions, and environmental conditions. Consequently, model selection should be based on application-specific experiments rather than assumptions derived from benchmark results alone.

Table 1. Comparison of object detection approaches for CCTV-based waste monitoring

Detection approach	Main characteristics	Potential application	Main limitation
Faster R-CNN	Two-stage detection	Detailed waste localization	Higher computational requirements
SSD	Single-stage detection	Lightweight visual monitoring	Small-object detection challenges
YOLOv4	Real-time detection	General waste identification	Requires application-specific adaptation
YOLOv7	Improved speed–accuracy balance	Multi-category waste detection	Computational requirements vary
YOLOv8	Anchor-free detection framework	Real-time residential surveillance	Requires locally representative training data

3.3. Dataset development and environmental adaptation

The availability of representative training data is essential for reliable waste detection. Existing waste image datasets, including TACO, provide examples of litter and discarded materials in different real-world environments [17]. Research on deep learning-based waste classification also demonstrates the potential of visual recognition for improving waste identification and recycling operations [18], [19]. However, image datasets designed for waste classification do not necessarily represent the temporal and spatial characteristics of illegal dumping captured by fixed CCTV cameras.

A residential CCTV dataset should include different camera heights, viewing angles, distances, and illumination conditions. It should also represent variations in waste appearance, including partially concealed objects, overlapping materials, and waste located near vegetation or roadside infrastructure. Training images should contain both positive and negative examples to reduce false detections. Most importantly, video sequences should be annotated with event-level information to distinguish newly deposited waste from previously existing garbage.

4. OBJECT TRACKING AND ILLEGAL DUMPING EVENT RECOGNITION

4.1. The role of multi-object tracking

Object tracking extends object detection by maintaining the identity of detected objects across consecutive video frames. This capability is important because illegal dumping is a temporal activity involving interactions between people, vehicles, and discarded materials. A tracking algorithm can associate a detected person with a persistent identity and estimate the person's movement through the monitored area. It can also maintain the trajectory of a vehicle or waste object when sufficient visual information is available [20], [21].

Several tracking algorithms have been developed to improve the continuity and reliability of object trajectories. SORT combines motion estimation and data association to provide efficient online tracking [20]. DeepSORT extends this approach by incorporating appearance information, which helps maintain object identities during temporary occlusions [21]. More recent methods, including ByteTrack, StrongSORT, and OC-SORT, provide alternative strategies for improving association under challenging tracking conditions [22]-[24].

4.2. Integrating detection and tracking

The combination of YOLOv8 and object tracking provides a more comprehensive framework than object detection alone. YOLOv8 identifies waste, people, and vehicles in individual frames, while the tracking module associates these detections across time. The resulting trajectories can provide information about the movement of individuals and the appearance of waste within a monitored location. This information can subsequently be evaluated by an event recognition module to determine whether a suspicious disposal activity has occurred.

A potential event recognition sequence begins when a person or vehicle enters a predefined monitoring region. The system then observes whether the individual carries or transports an object toward an unauthorized disposal location. A potential dumping event is identified when the object is deposited and remains in the monitored area after the person or vehicle moves away. The system should verify the appearance of new waste and the temporal relationship between the disposal activity and the observed objects before generating an incident notification.

Table 2 highlights that illegal dumping recognition requires cooperation between multiple processing components. The detector provides spatial information, while the tracker contributes temporal continuity. Event recognition then evaluates the relationships between these observations. This separation also allows each component to be improved independently without redesigning the entire monitoring architecture.

Table 2. Functional roles of detection, tracking, and event recognition

Processing component	Main function	Information generated
YOLOv8 detection	Identify people, vehicles, and waste	Object classes and bounding boxes
Multi-object tracking	Associate objects across frames	Track IDs and trajectories
Temporal analysis	Examine changes in scene conditions	Object appearance and disappearance
Event recognition	Identify potential disposal activities	Dumping event classification
Alert generation	Report verified incidents	Timestamp, location, and video evidence

4.3. Selecting an appropriate tracking algorithm

SORT offers a computationally efficient solution for applications with relatively stable camera views and limited occlusion. DeepSORT may be more appropriate when multiple individuals appear simultaneously or when people temporarily disappear behind vehicles and other obstacles. ByteTrack improves detection-to-track association by considering both high-confidence and lower-confidence detections, which may be useful when objects are partially obscured. StrongSORT and OC-SORT provide additional alternatives for applications requiring improved identity continuity and robust motion estimation [22]-[24].

The comparison in Table 3 indicates that no single tracking algorithm is universally optimal for residential waste monitoring. The selection should consider camera stability, crowd density, processing resources, and the duration of typical dumping activities. For a fixed residential CCTV installation, SORT or ByteTrack may offer practical starting points. More complex approaches should be considered when occlusion and repeated identity switching substantially affect event recognition.

Table 3. Comparison of object tracking algorithms

Component	Input	Output/function
SORT	Kalman filtering and Hungarian assignment	Low computational complexity
DeepSORT	Motion and appearance association	Improved identity continuity
ByteTrack	Association of high- and low-confidence detections	Improved detection utilization
StrongSORT	Enhanced appearance and motion association	More robust tracking
OC-SORT	Observation-centric motion association	Improved motion robustness

5. PROPOSED CCTV-BASED MONITORING FRAMEWORK

5.1. System architecture

An integrated monitoring framework can be organized into five interconnected processing stages: video acquisition, object detection, object tracking, dumping event recognition, and municipal notification. Existing CCTV cameras provide continuous video streams from selected residential locations. The video frames are subsequently processed by a YOLOv8 model to detect relevant objects. The detection results are transferred to a tracking module, which maintains object identities and trajectories across consecutive frames.

The event recognition component evaluates the temporal relationship between people, vehicles, and waste objects. It also compares the current scene with previously observed conditions to determine whether new waste has appeared. When the observed sequence satisfies predefined dumping criteria, the system generates an incident record. The record can be transmitted to a municipal dashboard containing the camera identification, location, timestamp, confidence score, and selected video evidence.

5.2. Temporal event recognition

Temporal event recognition is arguably the most important component of the proposed framework. A detection-only system may generate alerts whenever waste is visible, including locations where garbage is legitimately stored or awaiting scheduled collection. A tracking-based system can reduce this problem by analyzing whether waste appears during a sequence involving a person or vehicle. However, the system should not automatically interpret every person–waste interaction as illegal behavior.

A more reliable approach combines object tracking with background comparison and predefined spatial rules. For example, a new waste object appearing within a restricted area can trigger an initial event candidate. The system subsequently evaluates whether the object remains after the associated person or vehicle leaves the monitored region. Additional contextual information, including authorized collection schedules and designated waste disposal locations, can improve classification reliability. Human verification should remain available before enforcement actions are initiated.

5.3. Real-time processing and deployment

Real-time monitoring requires an appropriate balance between processing speed, detection quality, and system cost. Processing every frame at the original camera resolution may create unnecessary computational demands, particularly when multiple CCTV cameras operate simultaneously. Frame sampling, image resizing, and optimized inference can help reduce the processing workload. Edge computing can also reduce communication latency and limit the amount of video transmitted to centralized servers [25], [26].

The proposed framework can be deployed using existing CCTV infrastructure connected to a local edge computer or municipal server. A centralized dashboard can provide authorized personnel with access to incident records and monitoring statistics. The system should also support video retention policies and controlled access to recorded evidence. These arrangements are important because the proposed technology is intended to strengthen environmental management rather than introduce unrestricted surveillance of residential communities.

6. PERFORMANCE EVALUATION AND IMPLEMENTATION CHALLENGES

6.1. Detection and tracking performance

The effectiveness of an integrated monitoring system should be assessed using several complementary performance indicators. Object detection performance can be measured through precision, recall, F1-score, and mean average precision. Tracking performance can be assessed using multiple-object tracking accuracy, identity switches, and trajectory continuity. Real-time operation should additionally be evaluated through processing speed, inference latency, and computational resource utilization.

High detection accuracy does not necessarily indicate effective illegal dumping recognition. A system may accurately identify waste objects but still generate false alarms because it cannot distinguish unauthorized disposal from legitimate waste collection. Therefore, evaluation should include event-level precision, event-level recall, false alerts per camera per day, and notification delay. These measures are more directly connected to the practical requirements of municipal sanitation operations.

6.2. Environmental and technical challenges

Residential environments introduce several visual challenges that may reduce model performance. Poor nighttime illumination, rain, shadows, camera vibration, and partial occlusion can affect object localization and tracking continuity. Waste materials also exhibit substantial variation in size, shape, color, and texture, making reliable classification difficult. These conditions emphasize the importance of collecting representative local CCTV footage and evaluating the system under different environmental conditions [27], [28].

Another challenge concerns the relationship between visual evidence and legal definitions of illegal dumping. A computer vision system can identify suspicious disposal behavior, but it cannot independently establish ownership, authorization, or legal responsibility. Municipal regulations and designated collection schedules must therefore be incorporated into the decision-making process. Automated notifications should be treated as preliminary incident information requiring appropriate verification rather than definitive evidence of an offense.

6.3. Privacy and responsible deployment

The use of CCTV-based artificial intelligence in residential areas introduces important privacy and data governance considerations. Monitoring systems may capture identifiable individuals, vehicle registration numbers, and activities unrelated to waste disposal. Unnecessary identification should therefore be avoided when the environmental monitoring objective can be achieved through object detection and anonymized tracking. Privacy-preserving techniques, access controls, secure storage, and limited retention periods should be included in the system design [25], [29].

Municipal authorities should establish clear procedures governing the collection, processing, storage, and use of surveillance information. The monitoring framework should prioritize environmental incident detection rather than generalized behavioral surveillance. Automated alerts should be accessible only to authorized sanitation personnel and relevant administrative units. Such safeguards can improve public acceptance while maintaining the operational value of intelligent CCTV monitoring.

7. RESEARCH GAPS AND FUTURE DIRECTIONS

Although YOLO-based detection and multi-object tracking have advanced considerably, their integration for illegal waste dumping recognition remains a specialized research challenge. Existing waste detection research frequently emphasizes object classification, recycling, and waste sorting rather than the temporal recognition of unauthorized disposal activities [18], [19]. Similarly, conventional multi-object tracking benchmarks primarily evaluate pedestrian and vehicle tracking rather than interactions involving newly deposited waste. These differences indicate the need for dedicated datasets and evaluation protocols designed specifically for environmental surveillance.

Future research should prioritize the development of annotated residential CCTV video datasets containing normal activities, authorized collection events, and illegal dumping scenarios. Such datasets should include temporal event labels and contextual information about designated disposal locations. Another important direction involves combining YOLOv8 with lightweight tracking algorithms and background-change analysis to distinguish newly deposited waste from existing environmental clutter. This approach could improve incident recognition without requiring an excessively complex video understanding model.

The integration of computer vision with geographic information systems and municipal sanitation databases also offers considerable potential. Repeated detections can be aggregated to identify locations experiencing frequent dumping incidents. These spatial patterns may help authorities prioritize inspections, improve waste collection schedules, and evaluate sanitation interventions. In this context, intelligent surveillance can evolve from a reactive detection tool into a decision-support mechanism for urban environmental management.

8. CONCLUSION

CCTV-based illegal waste dumping detection offers a promising opportunity to connect advances in computer vision with practical urban sanitation management. YOLOv8 provides a suitable foundation for identifying waste, people, and vehicles in residential surveillance footage, while multi-object tracking contributes temporal information about their movements and interactions. However, reliable dumping recognition requires more than identifying garbage in individual video frames. It requires the integration of object detection, tracking, scene-change analysis, contextual rules, and temporal event recognition. The proposed framework emphasizes the distinction between detecting waste and recognizing unauthorized disposal activities. Its practical implementation should consider local environmental conditions,

representative datasets, real-time processing requirements, privacy protection, and municipal operational procedures.

Future research should prioritize dedicated CCTV video datasets and event-level evaluation to establish reliable performance under realistic residential conditions. Ultimately, integrating intelligent surveillance with municipal sanitation services could improve incident reporting, support targeted environmental interventions, and contribute to cleaner and healthier residential communities.

REFERENCES

[1] N. Ferronato and V. Torretta, "Waste Mismanagement in Developing Countries: A Review of Global Issues," *International Journal of Environmental Research and Public Health*, vol. 16, no. 6, p. 1060, Mar. 2019, doi: 10.3390/ijerph16061060.

[2] A. Siddiqua, J. N. Hahladakis, and W. A. K. A. Al-Attiya, "An overview of the environmental pollution and health effects associated with waste landfilling and open dumping," *Environmental Science and Pollution Research*, vol. 29, no. 39, pp. 58514–58536, Aug. 2022, doi: 10.1007/s11356-022-21578-z.

[3] S. Kaza, L. C. Yao, P. Bhada-Tata, F. Van Woerden, and D. Levine, *What a Waste 2.0: A Global Snapshot of Solid Waste Management to 2050*. Washington, DC: World Bank, 2018. doi: 10.1596/978-1-4648-1329-0.

[4] Y. Myagmar-Ochir and W. Kim, "A Survey of Video Surveillance Systems in Smart City," *Electronics (Basel).*, vol. 12, no. 17, p. 3567, Aug. 2023, doi: 10.3390/electronics12173567.

[5] E. Badidi, K. Moumane, and F. El Ghazi, "Opportunities, Applications, and Challenges of Edge-AI Enabled Video Analytics in Smart Cities: A Systematic Review," *IEEE Access*, vol. 11, pp. 80543–80572, 2023, doi: 10.1109/ACCESS.2023.3300658.

[6] 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, no. 11, pp. 3212–3232, Nov. 2019, doi: 10.1109/TNNLS.2018.2876865.

[7] L. Liu *et al.*, "Deep Learning for Generic Object Detection: A Survey," *International Journal of Computer Vision*, vol. 128, no. 2, pp. 261–318, Feb. 2020, doi: 10.1007/s11263-019-01247-4.

[8] J. Terven, D.-M. Córdova-Esparza, and J.-A. Romero-González, "A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS," *Mach. Learn. Knowl. Extr.*, vol. 5, no. 4, pp. 1680–1716, Nov. 2023, doi: 10.3390/make5040083.

[9] H. Abdu and M. H. Mohd Noor, "A Survey on Waste Detection and Classification Using Deep Learning," *IEEE Access*, vol. 10, pp. 128151–128165, 2022, doi: 10.1109/ACCESS.2022.3226682.

[10] S. Jin, Z. Yang, G. Królczykg, X. Liu, P. Gardoni, and Z. Li, "Garbage detection and classification using a new deep learning-based machine vision system as a tool for sustainable waste recycling," *Waste Management*, vol. 162, pp. 123–130, May 2023, doi: 10.1016/j.wasman.2023.02.014.

[11] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2016, pp. 779–788. doi: 10.1109/CVPR.2016.91.

[12] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017, doi: 10.1109/TPAMI.2016.2577031.

[13] W. Liu *et al.*, "SSD: Single Shot MultiBox Detector," In: *Leibe, B., Matas, J., Sebe, N., Welling, M. (eds) Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science*, vol 9905. Springer, Cham, 2016, pp. 21–37. doi: 10.1007/978-3-319-46448-0_2.

[14] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal Loss for Dense Object Detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020, doi: 10.1109/TPAMI.2018.2858826.

[15] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020.

[16] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2023, pp. 7464–7475. doi: 10.1109/CVPR52729.2023.00721.

[17] P. F. Proença and P. Simões, "TACO: Trash Annotations in Context Dataset," Jun. 2019, *Zenodo*. doi: 10.5281/zenodo.3354286.

[18] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," in *2016 IEEE International Conference on Image Processing (ICIP)*, IEEE, Sep. 2016, pp. 3464–3468. doi: 10.1109/ICIP.2016.7533003.

[19] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in *2017 IEEE International Conference on Image Processing (ICIP)*, IEEE, Sep. 2017, pp. 3645–3649. doi: 10.1109/ICIP.2017.8296962.

[20] Y. Zhang *et al.*, "ByteTrack: Multi-object Tracking by Associating Every Detection Box," 2022, pp. 1–21. doi: 10.1007/978-3-031-20047-2_1.

[21] Y. Du *et al.*, "StrongSORT: Make DeepSORT Great Again," *IEEE Trans. Multimedia*, vol. 25, pp. 8725–8737, 2023, doi: 10.1109/TMM.2023.3240881.

[22] J. Cao, J. Pang, X. Weng, R. Khirodkar, and K. Kitani, "Observation-Centric SORT: Rethinking SORT for Robust Multi-Object Tracking," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2023, pp. 9686–9696. doi: 10.1109/CVPR52729.2023.00934.

[23] G. Maggiolino, A. Ahmad, J. Cao, and K. Kitani, "Deep OC-Sort: Multi-Pedestrian Tracking by Adaptive Re-Identification," in *2023 IEEE International Conference on Image Processing (ICIP)*, IEEE, Oct. 2023, pp. 3025–3029. doi: 10.1109/ICIP49359.2023.10222576.

[24] T. I. Amosa *et al.*, "Multi-camera multi-object tracking: A review of current trends and future advances," *Neurocomputing*, vol. 552, p. 126558, Oct. 2023, doi: 10.1016/j.neucom.2023.126558.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[26] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, Jun. 2018, pp. 4510–4520. doi: 10.1109/CVPR.2018.00474.

[27] Y. Zhu *et al.*, "Improving Semantic Segmentation via Video Propagation and Label Relaxation," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2019, pp. 8848–8857. doi: 10.1109/CVPR.2019.00906.

[28] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and Efficient Object Detection," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2020, pp. 10778–10787. doi: 10.1109/CVPR42600.2020.01079.

[29] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for Object Detection," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jul. 2017, pp. 936–944. doi: 10.1109/CVPR.2017.106.