Journal of Sustainable Development of Energy, Water and Environment Systems

http://www.sdewes.org/jsdewes

Year 2026, Volume 14, Issue 2, 1130654

Original Research Article

Performance Assessment of You Only Look Once Models in Drone-Based Plastic Litter Quantification on Coastal Beaches

Mounir Grari[*1], *Mimoun Yandouzi*[2], *Abdelaadim Khriss*[3], *Aissa Kerkour Elmiad*[3], *Mohammed Badaoui*[4], *Alae-Eddine Barkaoui*[5], *Yassine Zarhloule*[5]

[1]Lab. MATSI, ESTO, Mohammed First University, Oujda, Morocco
[2]Lab. LSI, ENSAO, Mohammed First University, Oujda, Morocco
[3]Lab. LARI, FSO, Mohammed First University, Oujda, Morocco
[4]Lab. LaMSD, ESTO, Mohammed First University, Oujda, Morocco
[5]2GPMH Lab., FSO, Mohammed First University, Oujda, Morocco
e-mail: m.grari@ump.ac.ma

Cite as: Grari, M., Yandouzi, M., Khriss, A., Kerkour Elmiad, A., Badaoui, M., Barkaoui, A. E., Zarhloule, Y., Performance Assessment of You Only Look Once Models in Drone-Based Plastic Litter Quantification on Coastal Beaches, J.sustain. dev. energy water environ. syst., 14(2), 1130654, 2026, DOI: https://doi.org/10.13044/j.sdewes.d13.0654

ABSTRACT

Plastic pollution poses a growing threat to global ecosystems and human well-being, underscoring the urgent need for advanced methods that support accurate monitoring and effective mitigation efforts. Recent technological developments have transformed environmental assessment practices, particularly through the integration of aerial systems with artificial intelligence. The integration of unmanned aerial vehicles with deep learning techniques offers considerable advantages by enabling the coverage of large areas while providing rapid and precise analysis of collected data. The You Only Look Once family of deep learning models has gained prominence for real-time object detection, tracking, and counting, making it suitable for large-scale environmental applications. This study evaluates multiple versions of the You Only Look Once model to determine their effectiveness in quantifying plastic litter on coastal beaches using aerial imagery. The performance of versions 8, 9, 10, and 11 is evaluated in terms of detection, tracking, and counting accuracy using a custom dataset comprising short aerial video sequences. All evaluated versions demonstrated high counting performance, with accuracy values exceeding 98%. Version 11 achieved the strongest results overall, including the shortest inference times. These outcomes highlight the potential of advanced deep learning–based detection systems to enhance automated environmental monitoring and support more efficient management of plastic pollution in coastal environments.

KEYWORDS

Plastic litter, *Drones*, *Deep learning*, *You Only Look Once*, *Object tracking.*

INTRODUCTION

Plastic pollution has emerged as one of the most pressing environmental challenges globally, posing significant threats to ecosystems, biodiversity, and human health. The widespread adoption of plastic materials, driven by their convenience and low cost, has led to an unprecedented surge in production. According to the Organisation for Economic Cooperation and Development, global plastic production reached 460 million tons in 2019, with projections indicating that this figure could escalate to 975 million tons by 2050 if

[*] Corresponding author

current trends continue [1]. Despite this exponential growth, recycling rates remain alarmingly low, at approximately 9%. At the same time, the overwhelming majority of plastic waste accumulates in landfills or disperses into natural environments, particularly marine and coastal ecosystems, where it undergoes degradation into microplastics. These microscopic plastic fragments, typically less than 5 millimetres in diameter, are readily ingested by marine organisms across various trophic levels, from zooplankton to larger fish species. This bioaccumulation process introduces toxic substances into the aquatic food web, with potential ramifications extending to human health through the consumption of seafood [2]. The persistence and toxicological impact of microplastics, combined with their increasing difficulty in detection and removal as they fragment further, underscore the urgent need for effective monitoring and remediation strategies.

Addressing this multifaceted challenge requires comprehensive monitoring approaches that can identify pollution sources, quantify waste distribution, and inform targeted intervention measures. However, conventional monitoring methods, including manual beach surveys and satellite-based remote sensing, suffer from significant limitations in terms of spatial coverage, temporal resolution, cost-effectiveness, and detection accuracy, particularly for smaller debris items. To overcome these constraints, unmanned aerial vehicles equipped with high-resolution cameras have emerged as a powerful tool for environmental monitoring. These versatile platforms provide rapid, large-area surveys with unprecedented spatial precision and temporal flexibility at relatively low operational costs [3]. Their operational agility enables access to remote or hazardous locations, frequent revisits to monitor temporal changes in pollution patterns, and data acquisition at customisable altitudes and viewing angles. By providing real-time visual data on plastic waste distribution and accumulation patterns, drone-based monitoring systems can significantly enhance the efficiency of clean-up operations and inform evidence-based policy decisions regarding waste management and prevention strategies.

The transformative potential of drone technology for pollution monitoring has been substantially amplified by recent advances in artificial intelligence, particularly in the domains of computer vision and deep learning. These computational techniques enable automated analysis of aerial imagery, facilitating rapid and accurate detection, classification, and quantification of plastic debris without requiring extensive manual interpretation. Deep learning algorithms, trained on large annotated datasets, can learn to recognise the visual characteristics of plastic waste items even in complex environmental contexts characterised by variable lighting conditions, diverse background textures, and partial occlusions [4]. This automated processing capability dramatically accelerates data analysis workflows, transforming raw aerial imagery into actionable intelligence regarding pollution distribution and severity. Moreover, the integration of real-time data acquisition through drones with immediate processing through deep learning models creates opportunities for adaptive monitoring strategies, where survey patterns can be dynamically adjusted based on detected pollution hotspots. These integrated systems represent a cost-effective, scalable, and environmentally sustainable approach to persistent monitoring of plastic pollution in marine and coastal environments [5], offering significant improvements over traditional labour-intensive manual surveys and coarse-resolution satellite observations.

Building upon these technological advances, this study presents a systematic comparative evaluation of four iterations of the YOLO (You Only Look Once) deep learning architecture – specifically versions 8, 9, 10, and 11 – to determine their relative effectiveness for detecting, tracking, and quantifying plastic waste on beaches using aerial imagery acquired by drones. This comparative approach enables evidence-based selection of optimal algorithms for deployment in operational monitoring systems. The experimental methodology involves training each model variant on a comprehensive dataset of 4320 drone-captured images featuring diverse plastic waste items under varying environmental conditions. Model performance is then rigorously evaluated through testing on 150 standardised 10-second video sequences recorded at Saïdia Beach, Morocco, a coastal location at the mouth of the Moulouya River characterised by significant plastic waste accumulation. This site was selected for its ecological relevance and documented pollution challenges, providing a realistic test case for model validation. The study objectives are threefold: first, to compare model performance across multiple metrics including detection accuracy, counting precision, training efficiency, and inference speed; second, to identify optimal model-application pairings based on specific operational requirements and constraints; and third, to demonstrate the practical viability and operational advantages of integrating drone platforms with state-of-the-art deep learning models for automated, continuous monitoring of plastic pollution in coastal environments.

RELATED WORKS

The integration of artificial intelligence with environmental monitoring has catalysed significant advances in automated pollution detection and quantification. Deep learning, a branch of artificial intelligence leveraging multi-layered neural networks to extract hierarchical features and recognise complex patterns, has fundamentally transformed computer vision, speech recognition, and autonomous robotics [6]. Applied to environmental science, these techniques address the critical need for scalable, automated analysis of large datasets. Deep learning models have been increasingly deployed for detecting and classifying environmental contamination, including air pollutants, water quality indicators, and solid waste materials, demonstrating the potential to provide efficient, objective, and reproducible solutions for ecological monitoring while overcoming limitations of traditional manual approaches [7].

Central to environmental monitoring applications are three interconnected computer vision tasks: object detection, tracking, and counting. Object detection involves identifying and localising specific objects within images through bounding box coordinates, distinguishing it from simpler classification tasks. Modern detection systems utilise sophisticated machine learning models that can simultaneously process multiple object classes and generate spatial coordinates for each instance [8]. Building upon detection capabilities, object tracking maintains consistent identification of objects across video frames, essential for understanding dynamic behaviours. Object counting quantifies target objects, becoming challenging in crowded scenarios where instances overlap or appear at various scales. Traditional counting approaches include direct detection-based methods and density estimation techniques using convolutional neural networks to generate spatial density maps, particularly valuable when objects are densely packed [9]. State-of-the-art architectures have further advanced these techniques through the use of Fully Convolutional Networks and vision transformers, which estimate continuous density distributions. Meanwhile, attention mechanisms and self-supervised learning enhance robustness across diverse environmental conditions [10]. For video-based applications, Recurrent Neural Networks and Long Short-Term Memory networks provide temporal analysis capabilities, ensuring coherent tracking over extended periods [11]. These integrated capabilities enable the precise and automated quantification of plastic pollution in coastal zones, generating reliable data essential for scientific research and informed policy formulation.

Addressing the need for real-time detection capabilities, YOLO (You Only Look Once) has emerged as an influential family of object detection models. Introduced by Redmon *et al.* in 2016, YOLO pioneered a unified approach by reformulating detection as a single regression problem rather than a multi-stage pipeline [12]. Unlike region-based methods, which first generate potential regions and then classify them, YOLO processes entire images in one forward pass, directly predicting bounding box coordinates and class probabilities. This innovation dramatically accelerates inference speed, making YOLO well-suited for real-time processing applications. The architecture has evolved continuously through successive versions, each incorporating refinements to enhance accuracy, efficiency, and robustness. YOLOv8 (2020) introduced deeper convolutional architectures and feature pyramid networks for multi-scale detection, achieving a mean average precision of 83.4% on COCO [13]. YOLOv9 (2022) integrated knowledge distillation and quantisation methods with EfficientNet components, reaching 85.4% precision [14]. YOLOv10 (2022) emphasised speed optimisation through network pruning and Shuffle Net elements, achieving 84.2% accuracy while reducing computation time by 30% [15]. Most recently, YOLOv11 (2023) incorporated attention mechanisms and Vision Transformer principles, establishing a new benchmark with a mean average precision of 86.2% [16].

Translating these detection algorithms into practical environmental monitoring applications requires appropriate deployment platforms. Unmanned aerial vehicles equipped with high-resolution cameras and stabilised gimbals provide ideal platforms for applying computer vision algorithms to coastal monitoring challenges [17]. Their capacity to rapidly survey extensive areas while maintaining spatial resolution makes them invaluable for collecting visual data suitable for computational analysis. When integrated with YOLO algorithms, drone platforms create robust systems capable of autonomously identifying and quantifying plastic litter, enabling adaptive flight path planning that responds dynamically to detected pollution concentrations and optimising resource allocation for clean-up interventions.

Empirical validation of deep learning-drone integration for environmental monitoring has been demonstrated across multiple application domains. Nocua *et al.* [18] evaluated convolutional neural network architectures for urban traffic monitoring on resource-constrained embedded platforms, finding that MobileNet-v1-SSD achieved an optimal balance among accuracy (90%), recall (66%), and latency (10 milliseconds) for real-time applications. Their system incorporated Lucas-Kanade optical flow tracking with identity management, demonstrating the feasibility of deploying sophisticated computer vision on embedded GPU platforms. Directly applicable to plastic pollution monitoring, Maharjan *et al.* [19] developed an automated detection system for plastic waste in riverine environments using YOLO variants and drone imagery at two sites in Southeast Asia. Their comparative analysis revealed that YOLOv5s achieved a mean average precision of 0.81 without transfer learning, offering an excellent trade-off between accuracy and efficiency for resource-limited drones. At the same time, pre-trained YOLOv4 with fine-tuning attained the highest accuracy at 0.83. Notably, transfer learning substantially improved model generalisation when transferring knowledge between geographical sites, suggesting pre-trained models can accelerate deployment in new locations. Their investigation highlighted the critical influences of image resolution and plastic characteristics (type, shape, size, and colour) on detection performance.

Extending beyond detection to tracking applications, Shen *et al.* [20] investigated pollution source localisation using multi-UAV systems with target motion models, confirming tracking accuracy while identifying limitations of nadir-looking camera orientations that constrain observation capabilities. They recommended adjustable camera angles for improved tracking precision. In related work on aerial scene understanding, Rajagopal *et al.* [21] developed an optimised classification framework (ORNBFE) employing deep residual networks with automated hyperparameter tuning via Self-adaptive Global-best Harmony Search, achieving accuracies of 97.32% and 97.84% on the UCM and WHU-RS datasets, respectively. Collectively, these studies establish that integrating deep learning with drone platforms provides viable solutions for automated environmental monitoring, enabling scalable surveillance of plastic pollution across diverse geographical contexts.

MATERIALS AND METHODS

Automated plastic waste monitoring systems serve multiple stakeholders within environmental management and marine conservation. Coastal management authorities can deploy this technology for systematic beach monitoring, enabling evidence-based resource allocation to pollution hotspots. Environmental protection agencies benefit from objective data for regulatory compliance monitoring, while non-governmental organisations can utilise these systems for citizen science initiatives and advocacy. Research institutions gain access to high-resolution temporal datasets for investigating pollution dynamics and ecosystem impacts, while waste management companies can optimise collection routes based on real-time mapping. International bodies tracking sustainable development goals benefit from standardised, comparable data across regions. The economic efficiency of drone-based monitoring, compared to labour-intensive manual surveys, makes this approach particularly valuable for developing nations with limited monitoring budgets.

Implementing this monitoring capability requires an integrated methodological framework combining unmanned aerial vehicles with deep learning algorithms for automated detection, tracking, and quantification. The system architecture, illustrated in **Figure 1**, comprises three coordinated components. First, aerial data acquisition employs drones equipped with high-resolution cameras, conducting systematic surveys along pre-programmed flight paths to ensure complete spatial coverage. During operations, the drone continuously captures imagery while onboard detection systems, powered by trained YOLO models running on edge computing hardware or ground-based stations, identify plastic waste items. Upon detection, the system records GPS coordinates, creating georeferenced pollution maps, draws bounding boxes with associated confidence scores, and captures frames with metadata (including timestamp, altitude, camera angle, and detection confidence). Advanced configurations enable adaptive flight behaviour, adjusting altitude, zoom, or capturing multiple angles to improve classification confidence when environmental factors such as shadows, glare, or partial burial complicate identification.

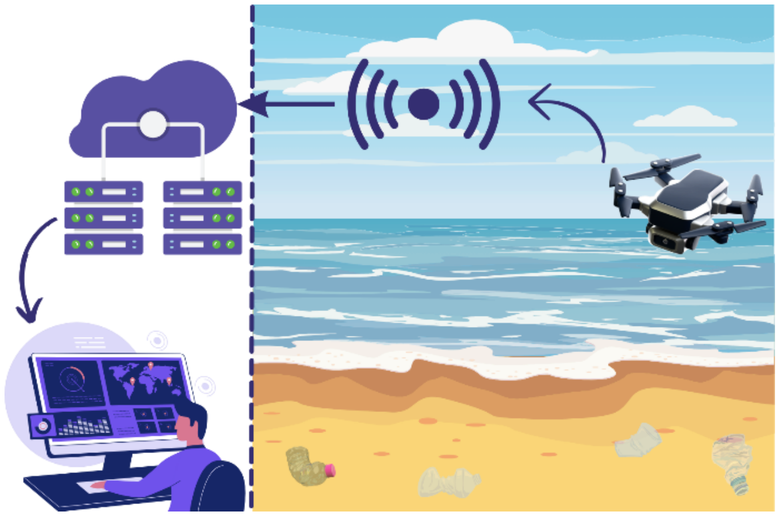

Figure 1. Proposed system architecture

The second component processes acquired imagery using deep learning models trained specifically for the recognition of coastal plastic waste. This phase executes three interconnected tasks: object detection identifies and localises plastic items; object tracking maintains consistent identification across video frames to prevent duplicate counting; and object counting aggregates detections while accounting for tracking information. The system employs algorithms that associate detections across frames based on spatial proximity, visual similarity, and motion patterns, maintaining unique identifiers for each tracked object. This ensures that a plastic bottle appearing in multiple consecutive frames is counted only once, rather than being counted repeatedly. **Figure 2** presents prototype testing of the plastic litter counting system, demonstrating the effectiveness of this integrated approach in real-world conditions.

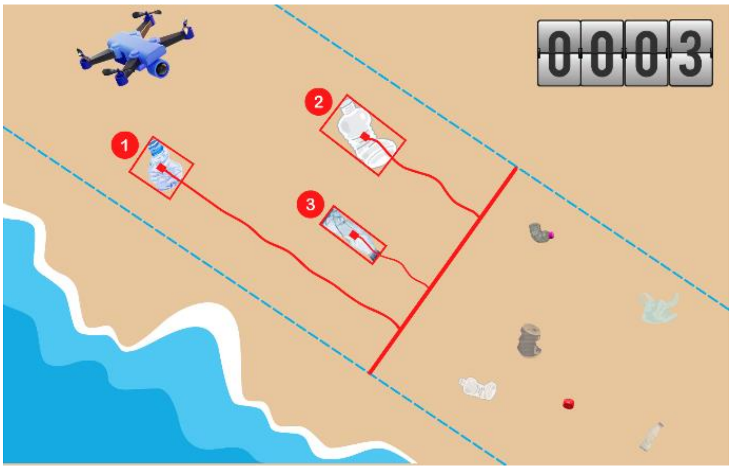

Figure 2. Prototype testing of a plastic litter counting system on beaches

The third component synthesises results into actionable intelligence, generating reports that include waste counts, spatial distribution maps, statistical summaries, and temporal trend analyses, all formatted for various stakeholder needs. The complete methodological workflow, illustrated in **Figure 3**, follows a structured sequence of operations designed to ensure reproducibility, accuracy, and operational efficiency, from initial data collection through final reporting. The specific procedures employed at each stage are detailed in the following subsections.

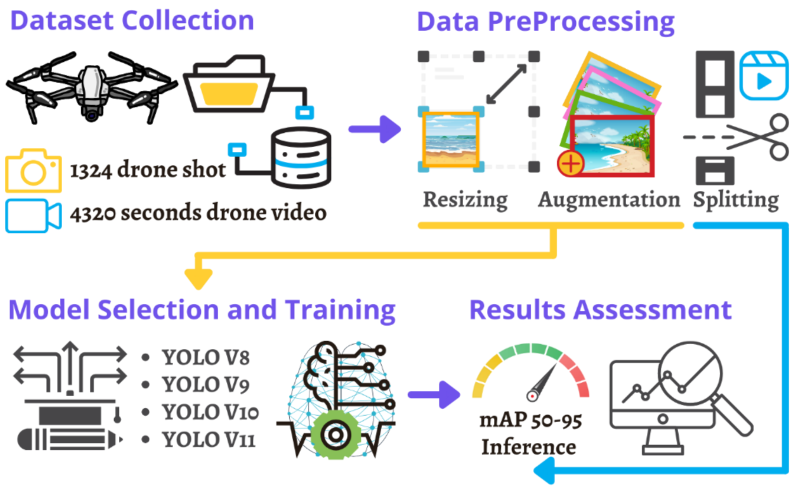

Figure 3. Research methodology

Dataset Collection

Data collection involved acquiring comprehensive imagery from beach areas containing both plastic waste and debris-free zones, captured by drones at varying altitudes and angles. The study site, situated at the mouth of the Moulouya River in eastern Morocco, was selected for its ecological diversity, encompassing forests, steppes, and wetlands, which provide a representative dataset for training detection models and contribute to an understanding of Mediterranean coastal pollution. Through multiple site visits across different seasons, 4320 images of plastic waste were collected. Additionally, 1324 seconds of drone footage were recorded to create a video database for evaluating model performance in plastic waste counting. This video collection enables the assessment of the model's ability to track and count plastic waste in dynamic, real-time environments, thereby enhancing the robustness of the detection system and providing comprehensive testing resources for realistic scenarios.

Data Pre-Processing

Following data collection, pre-processing ensured dataset quality and consistency through data refinement operations, including intensity normalisation, distortion correction, and image scaling, which produced standardised input facilitating model convergence. Data augmentation introduced variations (rotations, flips, zooms) to enhance dataset diversity and improve generalisation across real-world scenarios. The collected drone videos were split and normalised into 150 ten-second sequences, each manually labelled with plastic waste item counts to provide a reliable ground truth for model evaluation. This annotated dataset serves as the benchmark for final testing, ensuring rigorous assessment of detection, tracking, and counting capabilities in real-world conditions.

Model Selection and Training Process

The comparative evaluation focuses on four YOLO model variants, YOLOv8, YOLOv9, YOLOv10, and YOLOv11, selected for their optimised architectures and task adaptability. The training phase involves meticulous fine-tuning and optimisation to leverage available data and enhance performance. Following pre-processing and augmentation, the dataset was divided into training (70%), validation (20%), and testing (10%) subsets, enabling robust learning and unbiased performance evaluation. During training, each object was precisely outlined with bounding boxes and categorised using open-source annotation software, ensuring accurate identification and distinction of plastic waste. The training pipeline aligns with model-specific requirements, incorporating TXT-based annotations and YAML configuration files to optimise learning efficiency and effectiveness. This phase aims to develop models capable of highly accurate plastic waste detection, tracking, and counting, thereby providing stakeholders with effective solutions for monitoring and mitigating the impact of plastic pollution.

RESULTS

To rigorously assess the performance of the developed models, a thorough analysis of the collected data was conducted. This evaluation phase involved allocating 10% of the test set to assess both the mean precision and inference speed of each model variant. In addition to testing on pre-processed image data, the previously constructed database of 150 drone video sequences, each lasting 10 s and specifically designed for counting simulations, was employed for validation. These videos, manually labelled with the exact number of plastic items present, enabled the determination of the final accuracy of the models using the accuracy metric, which proved most suitable for this application. This validation stage provides a clear understanding of model performance on unseen data, highlighting the ability to detect, track, and quantify plastic waste while assessing potential for real-time deployment via drones. To ensure a comprehensive evaluation, multiple performance metrics were utilised, allowing for a detailed and multidimensional analysis of model effectiveness. The following subsections present the hardware specifications employed, the evaluation metrics utilised, and the experimental results obtained from comparative testing of the model variants.

Hardware Specifications

The experimental infrastructure comprised both aerial data acquisition equipment and computational resources for model training and evaluation. For aerial data collection, a DJI Mavic Air unmanned aerial vehicle equipped with a high-resolution camera was used, ensuring detailed and accurate image capture for object detection, tracking, and counting applications [22]. The computational infrastructure for training and testing object-detection models consisted of a Dell PowerEdge R740 server equipped with an Intel Xeon Silver 4210 processor operating at 2.2 GHz and 80 GB of random-access memory. Additionally, this server was configured with two NVIDIA RTX A5000 graphics processing units, each providing 24 GB of dedicated graphics memory to accelerate deep learning computations.

Evaluation Metrics

Various quantitative metrics exist for assessing the performance of object detection algorithms. The evaluation framework employed in this study incorporated multiple complementary metrics to provide a comprehensive assessment of model capabilities. Mean Average Precision represents an essential metric for evaluating model performance, especially in standardised benchmarks. This metric is calculated as the mean of Average Precision values across varying thresholds, typically ranging from 50% to 95% in 5% increments. The Average Precision for a given class is derived from the Precision-Recall curve according to the following formulation:

$$AP = \sum_{k=0}^{k=n-1} [Recalls(k) - Recalls(k+1) \times Precisions(k)] \tag{1}$$

where *Recalls(n)*=0, *Precisions(n)*=1, and *n* denotes number of thresholds.

The mean Average Precision *(mAP)* is a comprehensive metric that evaluates model performance by considering precision and recall across different categories and thresholds, offering a detailed assessment of its effectiveness and robustness:

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{2}$$

For *mAP* in the range of 50–95%, the metric is computed as:

$$mAP_{50-95} = \frac{1}{10}\sum_{IoU=0.50}^{0.95} AP(IoU) \tag{3}$$

Intersection over Union *(IoU)* measures the overlap between a detected object and its ground truth, assessing localisation accuracy by comparing the intersection and union areas of two rectangles. While metrics like Average Precision *(AP)* and mean Average Precision *(mAP)* evaluate overall model performance, *IoU* specifically highlights the model's ability to localise objects accurately.

$$IoU = \frac{Area\ of\ Intersection}{Area\ of\ Union} \tag{4}$$

In the context of object counting applications, the accuracy metric evaluates the precision of model predictions in terms of correctly enumerating objects. This metric compares the predicted number of objects with the actual count, assessing how effectively the model identifies and counts objects. For this investigation, the effectiveness of different model versions was assessed using a test dataset comprising 150 video sequences, each lasting 10 s. The accuracy metric serves as the primary evaluation criterion, as it provides a direct measure of model capability to estimate the number of objects present in each sequence correctly. The accuracy for object counting is computed according to the following formula:

$$Accuracy = 1 - \frac{|\,Count_{predicted} - Count_{actual}\,|}{Count_{actual}} \quad (5)$$

where $Count_{predicted}$ is the number of objects detected by the model, and $Count_{actual}$ is the ground truth count of objects in the video sequence.

Inference time constitutes another critical performance indicator for deep learning models, referring to the duration required for the model to process an input and generate predictions. This metric is influenced by factors such as model architecture complexity and available computational resources, which directly affect performance in real-time applications. The objective in model optimisation involves reducing inference time without sacrificing detection accuracy, especially in scenarios requiring rapid response capabilities.

Comparative Model Performance

The quantitative performance evaluation of the four model variants across multiple metrics is presented in this subsection. All models were trained for seventy epochs to ensure convergence and optimise performance. **Table 1** presents the comprehensive results obtained from testing on the reserved test set, while **Figure 4** illustrates the temporal evolution of mean Average Precision at the 0.5 threshold across training epochs for the evaluated models on the validation dataset.

Table 1. Achieved results for the implemented models (on the testing set)

Model	Precision	Recall	*mAP@* 0.5	*mAP@* 0.95	*mAP* 50–95	Accuracy (counting)	Inference [s/image]
YOLOv8	98.47	98.22	96.05	89.14	92.61	99.02	~0.0022
YOLOv9	94.12	95.04	94.06	87.01	90.57	98.08	~0.0017
YOLOv10	96.71	96.15	95.97	89.09	92.58	98.45	~0.0031
YOLOv11	98.63	98.51	97.13	90.02	93.53	99.07	~0.0020

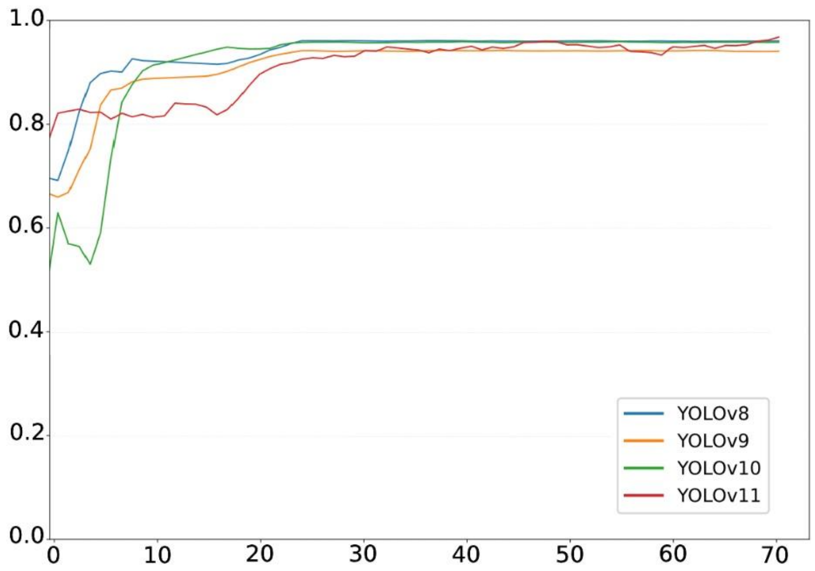

Figure 4. Evolution of mAP@0.5 with the number of epochs
for the evaluated models on the validation dataset

As demonstrated, the experimental results encompass various performance metrics, including precision, recall, mean Average Precision at confidence thresholds of 0.5 and 0.95, mean Average Precision across the 50–95% range, counting accuracy, and inference time for each model variant. All four model variants demonstrated the capability to detect, track, and count plastic litter, with performance ranging from good to excellent. The results reveal notable performance differences among the models that influence optimal selection depending on specific application priorities and operational constraints.

Notably, YOLOv8 and YOLOv11 exhibited significantly faster convergence during the training phase compared to the other models, requiring fewer than 30 epochs to reach optimal performance. In contrast, YOLOv9 and YOLOv10 required additional epochs to achieve comparable performance levels. Regarding detection accuracy metrics, YOLOv8 and YOLOv11 achieved mean Average Precision values of 92.61% and 93.53% respectively, across the 50–95% threshold range, substantially exceeding the performance of YOLOv9 (90.57%) and YOLOv10 (92.58%). However, when considering computational efficiency, YOLOv9 demonstrated the fastest inference time at approximately 0.0017 s per image, compared to 0.0022 s for YOLOv8, 0.0020 s for YOLOv11, and 0.0031 s for YOLOv10.

The video-based counting evaluation conducted on 150 ten-second sequences revealed that YOLOv8 and YOLOv11 achieved the highest counting accuracies of 99.02% and 99.07% respectively, outperforming YOLOv10 (98.45%) and YOLOv9 (98.08%). These results demonstrate the models' capability to maintain consistent detection and tracking performance across temporal sequences. **Figure 5** presents representative frames extracted from the counting process performed by the YOLOv8 model on drone-recorded video footage from the Saïdia Beach study area, illustrating the detection and tracking capabilities in realistic operational conditions.

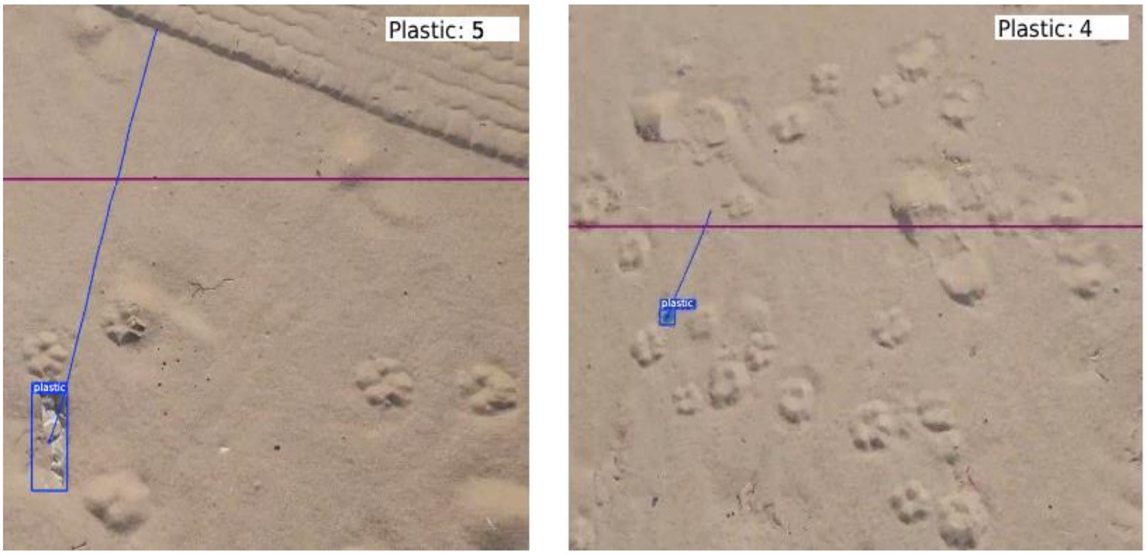

Figure 5. Sample frames of the counting process using
YOLOv8 on drone footage from Saïdia Beach

DISCUSSION

The experimental results provide important insights into the performance characteristics and operational trade-offs among the evaluated model variants for detecting and quantifying plastic waste in coastal environments. The comparative analysis demonstrates that while all four models achieve satisfactory performance for the intended application, distinct differences emerge when examining specific performance dimensions and deployment scenarios.

The superior mean Average Precision scores achieved by YOLOv8 (92.61%) and YOLOv11 (93.53%) across the threshold range of 50–95% indicate enhanced robustness in object localisation accuracy across varying degrees of spatial overlap between predicted and ground truth bounding boxes. This characteristic proves particularly valuable in complex environmental contexts where plastic waste items may be partially obscured by sand, vegetation, or other debris, or when objects appear at oblique viewing angles that complicate precise boundary delineation. The ability to maintain high detection confidence across stringent Intersection over Union thresholds minimises both false positive and false negative detections, thereby improving the reliability of automated counting systems. For applications prioritising accurate quantification over processing speed, such as systematic beach monitoring surveys conducted for scientific research or regulatory compliance assessment, these models represent optimal choices.

The rapid convergence behaviour exhibited by YOLOv8 and YOLOv11, reaching optimal performance within 30 training epochs compared to the extended training periods required by YOLOv9 and YOLOv10, suggests more efficient parameter optimisation and potentially superior architectural design for the specific characteristics of plastic waste detection tasks. This training efficiency translates to practical advantages, including reduced computational resource consumption during model development, faster iterative refinement when incorporating new training data, and a decrease in time-to-deployment for operational systems. The convergence patterns may reflect architectural innovations in these models that better capture the visual features distinguishing plastic waste from natural beach materials.

Conversely, YOLOv9 achieves the fastest inference time at 0.0017 s per image, representing approximately 23% faster processing than YOLOv8 and 15% faster than YOLOv11. This computational efficiency advantage positions YOLOv9 as potentially preferable for deployment scenarios with stringent real-time processing constraints, such as onboard processing on resource-limited embedded computing platforms mounted directly on unmanned aerial vehicles, or applications requiring immediate detection feedback to enable reactive flight path adjustment. However, this speed advantage necessitates accepting a modest reduction in detection accuracy (90.57% mean Average Precision), representing an approximate 2–3% decrease relative to the highest-performing models. The suitability of this accuracy-speed trade-off depends on specific application requirements and tolerance for counting errors.

The video-based counting evaluation provides particularly relevant validation for real-world deployment scenarios, as it assesses not only frame-by-frame detection capability but also the temporal consistency of tracking algorithms that prevent duplicate counting of identical objects appearing across multiple frames. The exceptional counting accuracies achieved by YOLOv8 (99.02%) and YOLOv11 (99.07%) demonstrate effective integration of detection and tracking components, ensuring that individual plastic items maintain consistent identities throughout their appearance in video sequences. This temporal coherence proves essential for generating reliable quantitative estimates from drone surveys, as even high per-frame detection accuracy becomes compromised if tracking failures cause the same object to be counted multiple times or lost between frames. The slightly lower counting accuracies of YOLOv10 (98.45%) and YOLOv9 (98.08%), while still representing strong overall performance, indicate greater susceptibility to tracking errors or detection inconsistencies across temporal sequences.

The demonstrated performance levels across all evaluated models confirm the technical feasibility of drone-based automated plastic pollution monitoring using contemporary deep learning architectures. The counting accuracies exceeding 98% for all variants indicate operational readiness for deployment in real-world environmental management applications, where such precision levels provide actionable quantitative data for decision-making regarding clean-up resource allocation and pollution trend assessment. The integration of aerial mobility, high-resolution imaging, and artificial intelligence creates a powerful monitoring capability that substantially exceeds traditional manual survey methods in terms of spatial coverage, temporal frequency, cost-effectiveness, and data reproducibility.

CONCLUSION

This study demonstrates that all four YOLO models – YOLOv8, YOLOv9, YOLOv10, and YOLOv11 – are capable of detecting, tracking, and counting plastic litter on beaches, with varying degrees of performance. Among them, YOLOv8 and YOLOv11 exhibit the best balance between accuracy, inference speed, and detection performance, making them particularly well-suited for quantifying plastic litter from drones. The *mAP*@50–95 results confirm their superiority in detection, achieving 92.61% and 93.53%, respectively, surpassing YOLOv9 and YOLOv10. Additionally, their rapid convergence during training (less than 30 epochs) highlights their efficiency in learning from data, enabling quicker model optimisation compared to YOLOv9 and YOLOv10.

Inference time is another critical factor, with YOLOv9 proving to be the fastest model (~0.0017s per image). However, this advantage comes at the cost of slightly lower detection accuracy (*mAP*@50–95 of 90.57%), which could affect applications where high precision is required. While this model may be preferable in real-time embedded applications where speed is paramount, its trade-off in precision limits its suitability for scenarios that demand accurate counting.

Accuracy in tracking and counting, tested over 150 manually labelled 10-second drone video sequences, reveals that YOLOv8 (99.02%) and YOLOv11 (99.07%) outperform the other models. Their ability to maintain precise detection over time ensures reliable quantification of plastic litter, which is crucial for environmental monitoring efforts. These results confirm that real-time drone-based detection using deep learning is a viable approach for assessing plastic pollution in coastal areas.

While these findings are promising, further research is needed to enhance model generalisation to diverse environmental conditions by incorporating larger and more varied datasets. Additionally, exploring hybrid approaches that combine YOLO models with transformer-based architectures could improve robustness against challenging detection scenarios, such as occlusions and varying lighting conditions. Finally, investigations of real-world deployment on autonomous drones with onboard processing will be conducted to assess the feasibility of real-time plastic litter monitoring in operational conditions.

ACKNOWLEDGEMENTS

We would like to extend our sincere appreciation to Mohammed First University for their financial and logistical support throughout the REMEDIES project. Our heartfelt thanks also go to the project contributors for their generous funding and support through the HORIZON Innovation Actions program, which aims to prevent and eliminate pollution in our oceans, seas, and waters (HORIZON-MISS-2021-OCEAN-03). We greatly appreciate the dedicated efforts of all project partners, researchers, and collaborators who have worked together to develop and implement effective strategies for both valorising and preventing plastic litter. The support of the PhD Association Scholarship-PASS offered by the National Centre of Scientific and Technical Research (CNRST), Morocco, with grant 77 UMP2023, is also recognised by A. Khriss.

REFERENCES

1. OECD Environment Statistics (Dataset) OCDE (2024), Global Plastics Outlook : Plastics use by region - projections , https://www.oecd.org/en/publications/global-plastics-outlook_de747aef-en.html, [Accessed: Nov.16, 2025].
2. Lorenzo-Navarro J, Castrillón-Santana M, Sánchez-Nielsen E, et al (2021) Deep learning approach for automatic microplastics counting and classification. Sci Total Environ 765:142728. https://doi.org/10.1016/J.SCITOTENV.2020.142728.

3. Topouzelis K, Papakonstantinou A, Garaba SP (2019) Detection of floating plastics from satellite and unmanned aerial systems (Plastic Litter Project 2018). Int J Appl Earth Obs Geoinf 79:175–183. https://doi.org/10.1016/J.JAG.2019.03.011.
4. Delgado, J., Barkaoui, A., Petelin, M., et al., *AI-Driven Aerial Drones and Monitoring App: New Developments to Facilitate Citizen Science Initiatives on Plastic Pollution Monitoring and Clean-Ups on Beaches*, Proceedings of the EGU General Assembly 2024, Vienna, Austria, 14–19 April 2024, EGU24-7368, https://doi.org/10.5194/egusphere-egu24-7368.
5. Fakayode SO, Mehari TF, Fernand Narcisse VE, et al (2024) Microplastics: Challenges, toxicity, spectroscopic and real-time detection methods. Appl Spectrosc Rev. https://doi.org/10.1080/05704928.2024.2311130.
6. Berrahal M, Boukabous M, Yandouzi M, et al (2023) Investigating the effectiveness of deep learning approaches for deep fake detection. Bull Electr Eng Informatics 12:3853–3860. https://doi.org/10.11591/EEI.V12I6.6221.
7. Zhu L, Husny ZJBM, Samsudin NA, et al (2023) Deep learning method for minimizing water pollution and air pollution in urban environment. Urban Clim 49:101486. https://doi.org/10.1016/J.UCLIM.2023.101486.
8. Liang, H., Song, H., Li, H. and Dai, Z., Vehicle Counting System Using Deep Learning and Multi-Object Tracking Methods, *Transportation Research Record*, Vol. 2674, No. 4, pp. 114-128, 2020, https://doi.org/10.1177/0361198120912742.
9. Grari M, Yandouzi M, Mohammed B, et al (2024) Comparative study of teachable machine for forest fire and smoke detection by drone. Bull Electr Eng Informatics 13:1970–1979. https://doi.org/10.11591/EEI.V13I3.6578.
10. Abuaiadah, D., Switzer, A., Bosu, M. and Liu, Y., Automatic Counting of Chickens in Confined Area Using the LCFCN Algorithm, *Proceedings of the 2022 International Conference on Intelligent Systems and Computer Vision (ISCV)*, Fez, Morocco, pp. 1–7, 2022, IEEE, https://doi.org/10.1109/ISCV54655.2022.9806092.
11. Guicquero, W., Sicard, G. and others, BILLNET: A Binarized Conv3D-LSTM Network with Logic-Gated Residual Architecture for Hardware-Efficient Video Inference, *Proceedings of the 2022 IEEE Workshop on Signal Processing Systems (SiPS)*, pp. 1–6, 2022, IEEE, https://doi.org/10.1109/SiPS55645.2022.9919206.
12. Redmon JSDRGAF (2016) (YOLO) You Only Look Once. Cvpr 2016-December:779–788. https://doi.org/10.1109/CVPR.2016.91.
13. Yandouzi M, Grari M, Berrahal M, et al (2023) Investigation of Combining Deep Learning Object Recognition with Drones for Forest Fire Detection and Monitoring. Int J Adv Comput Sci Appl 14:377–384. https://doi.org/10.14569/IJACSA.2023.0140342.
14. Wajdi AZ, Wibowo P, Sasaki S (2024) YOLOv8 vs. YOLOv9: Evaluating Object Detection Algorithms for Marine Waste Recognition. Dig Tech Pap - IEEE Int Conf Consum Electron 286–292. https://doi.org/10.1109/ISCT62336.2024.10791223.
15. Huynh TT, Nguyen HT, Phu DT (2024) Enhancing Fire Detection Performance Based on Fine-Tuned YOLOv10. Comput Mater Contin 81:2281–2298. https://doi.org/10.32604/cmc.2024.057954.
16. Alkhammash EH (2025) A Comparative Analysis of YOLOv9, YOLOv10, YOLOv11 for Smoke and Fire Detection. Fire 8:26. https://doi.org/10.3390/fire8010026.
17. Yandouzi, M., Grari, M., Berrahal, M., Kerkour, A., Elmiad, M. B., Barkaoui, A. and Zarhloule, Y., Elevating Plastic Litter Surveillance: Drones and Deep Learning for Efficient Detection, *Journal of Theoretical and Applied Information Technology*, Vol. 102, No. 11, pp. 4735–4751, 2024
18. Nocua M F, Pérez-Holguín WJ, Pardo-Beainy C (2025) Urban traffic monitoring based on deep learning on an embedded GPU. Expert Syst Appl 273:126847. https://doi.org/10.1016/j.eswa.2025.126847.

19. Maharjan N, Miyazaki H, Pati BM, et al (2022) Detection of River Plastic Using UAV Sensor Data and Deep Learning. Remote Sens 14:3049. https://doi.org/10.3390/rs14133049.
20. Shen B, Jiang J, Li D, Qian F (2023) Location and tracking of environmental pollution sources under multi-UAV vision based on target motion model. Front Ecol Evol 11:1178990. https://doi.org/10.3389/FEVO.2023.1178990/BIBTEX.
21. Rajagopal, A., Ramachandran, A., Shankar, K., Khari, M., Jha, S., Lee, Y. and Joshi, G. P., Fine-Tuned Residual Network-Based Features with Latent Variable Support Vector Machine-Based Optimal Scene Classification Model for Unmanned Aerial Vehicles, IEEE Access, Vol. 8, pp. 118396–118404, 2020, https://doi.org/10.1109/ACCESS.2020.3004233.
22. Support for Mavic Air – DJI, https://www.dji.com/global/support/product/mavic-air, [Accessed: Feb.07, 2025].

Paper submitted: 11.08.2025
Paper revised: 19.11.2025
Paper accepted: 25.11.2025