A project report on

DETECTION and MONITOTRING Of BURIED PLASTIC IN SANDY ENVIROMENTS USING COMPUTER VISION

Submitted in partial fulfillment for the award of the degree of

M.Tech. (Integrated) Computer Science and Engineering with Specialization in Business Analytics

By

GUNA SHANKAR (20MIA1162)

VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

CHENNAI

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

November, 2024

DETECTION and MONITOTRING Of BURIED PLASTIC IN SANDY ENVIROMENTS USING COMPUTER VISION

Submitted in partial fulfillment for the award of the degree of

M.Tech. (Integrated) Computer Science and Engineering with Specialization in Business Analytics

by

GUNA SHANKAR S (20MIA1162)

VIT®
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)
CHENNAI

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

November, 2024

VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

CHENNAI

School of Computer Science and Engineering

DECLARATION

I here by declare that the thesis entitled " **DETECTION and MONITOTRING Of BURIED PLASTIC IN SANDY ENVIROMENTS USING COMPUTER VISION** " submitted by me, for the award of the degree of M.Tech. (Integrated) Computer Science and Engineering with Specialization in Business Analytics, Vellore Institute of Technology, Chennai, is are cord of bonafide work carried out by me under the supervision of "**Dr. P SUBBULAKSHMI**".

I further declare that the work reported in this thesis has not been submitted and will not be submitted, either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university.

Place: Chennai

Date: 13/11/201 **Signature of the Candidate**

VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

CHENNAI

School of Computer Science and Engineering

CERTIFICATE

This is to certify that the report entitled **"DETECTION and MONITOTRING Of BURIED PLASTIC IN SANDY ENVIROMENTS USING COMPUTER VISION"** is prepared and submitted by **GUNA SHANKAR S (20MIA1162)** to Vellore Institute of Technology, Chennai, in partial fulfillment of the requirement for the award of the degree of **M.Tech. (Integrated) Computer Science and Engineering with Specialization in Business Analytics** programme is a bonafide record carried out under my guidance. The project fulfills the requirements as per the regulations of this University and in my opinion meets the necessary standards for submission. The contents of this report have not been submitted and will not be submitted either in part or in full, for the award of any other degree or diploma and the same is certified.

Signature of the Guide:

Name: Dr./Prof. **P. SUBBULAKSHMI**

Date:

Signature of the Examiner 1	Signature of the Examiner 2
Name: **Dr. Anita X**	Name: **Dr. Sobitha Ahila S**
Date:	Date:

Approved by the Head of Department

ABSTRACT

Plastic pollution represents one of the most critical challenges facing coastal ecosystems today. In these sensitive environments, where wind, tides, and human activities constantly reshape the landscape, plastics often become buried or partially submerged in sand, going undetected by conventional methods. This hidden waste poses severe risks to marine ecosystems, endangers wildlife, and threatens human health as it gradually degrades into microplastics, which can enter the food chain. Traditional techniques for detecting plastic pollution have limitations in accuracy, speed, and scope, often failing to recognize submerged or camouflaged plastic debris.

This research pioneers a novel approach that integrates high-resolution drone imagery with advanced deep learning and computer vision techniques to create a more precise, scalable, and automated system for detecting plastic pollution. Drones equipped with high-resolution cameras capture extensive visual data over sandy and coastal regions, providing comprehensive spatial coverage that enables real-time monitoring. The high-quality images gathered are processed using YOLO (You Only Look Once), a deep learning model designed for object detection and classification. YOLO is particularly advantageous in this context for its efficiency in detecting multiple objects in a single frame and for maintaining a high degree of accuracy, even with partially visible or submerged items.

Beyond the technical implementation, this research underscores the broader environmental and societal implications of the findings. By deploying a highly accurate and efficient system for plastic detection, stakeholders can gain invaluable insights into the distribution, concentration, and movement of plastic debris along coastlines. The system's data can support the design of targeted cleanup initiatives, the establishment of conservation policies, and the development of educational programs that highlight the urgency of plastic pollution. Additionally, this research has the potential to inform global efforts by providing an adaptable, scalable model that can be replicated in diverse coastal settings worldwide.

In summary, this project represents a significant advancement in environmental monitoring by combining drone technology, remote sensing, and deep learning for a comprehensive solution to plastic waste detection. The ability to accurately identify, classify, and track plastic pollution in sandy environments opens new avenues for research, cleanup, and prevention, providing a valuable resource in the global fight against plastic pollution in fragile coastal ecosystems.

ACKNOWLEDGEMENT

It is my pleasure to express with deep sense of gratitude to **Dr. P. SUBBULAKSHMI**,Associate Professor, School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, for her constant guidance, continual encouragement, understanding; more than all, she taught me patience in my endeavor. My association with her is not confined to academics only, but it is a great opportunity on my part of work with an intellectual and expert in the field of Computer Vision.

It is with gratitude that I would like to extend my thanks to the visionary leader Dr. G. Viswanathan our Honorable Chancellor, Mr. Sankar Viswanathan, Dr. Sekar Viswanathan, Dr. G V Selvam Vice Presidents, Dr. Sandhya Pentareddy, Executive Director, Ms. Kadhambari S. Viswanathan, Assistant Vice-President, Dr. V. S. Kanchana Bhaaskaran Vice-Chancellor, Dr.T. Thyagarajan Pro-Vice Chancellor, VIT Chennai and Dr. P. K. Manoharan, Additional Registrar for providing an exceptional working environment and inspiring all of us during the tenure of the course.

Special mention to Dr. Ganesan R, Dean, Dr. Parvathi R, Associate Dean Academics, Dr. Geetha S, Associate Dean Research, School of Computer Science and Engineering, Vellore Institute of Technology, Chennai for spending their valuable time and efforts in sharing their knowledge and for helping us in every aspect.

In jubilant state, I express ingeniously my whole-hearted thanks to Dr.Sivabalakrishnan. M, Head of the Department, Project Coordinator, Dr. Yogesh C, SCOPE, Vellore Institute of Technology, Chennai, for their valuable support and encouragement to take up and complete the thesis.

My sincere thanks to all the faculties and staff at Vellore Institute of Technology, Chennai, who helped me acquire the requisite knowledge. I would like to thank my parents for their support. It is indeed a pleasure to thank my friends who encouraged me to take up and complete this task.

Place: Chennai

Date: 13/11/2024 **Name of the student: GUNA SHANKAR S**

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

YOLO	YOU ONLY LOOK ONCE
GPR	GROUND-PENETRATING RADAR
CNNS	CONVOLUTIONAL NEURAL NETWORKS (CNNS)
MIDAS	MUNICH IMAGE DATA ANALYSIS SYSTEM (MIDAS)
ESO	EUROPEAN SOUTHERN OBSERVATORY (ESO)
ORB-SLAM3	ORIENTED FAST AND ROTATED BRIEF SIMULTANEOUS LOCALIZATION AND MAPPING
AR	AUGMENTED REALITY (AR)
ROIS	REGIONS OF INTEREST (ROIS)
FC	FULLY CONNECTED (FC)
LIDAR	LIGHT DETECTION AND RANGING (LIDAR)

CHAPTER 1

INTRODUCTION

1.1 THE GLOBAL PLASTIC WASTE CRISIS IN COASTAL ENVIRONMENTS

1.1.1 The Surge in Plastic Production and Its Environmental Impact

The exponential increase in plastic production globally has become a major environmental concern, with plastic waste now a ubiquitous pollutant affecting various ecosystems. Since the mid-20th century, the production of plastic has surged, with plastics being used in countless applications due to their versatility, durability, and low cost. However, these very characteristics make plastic a persistent environmental pollutant. Plastics do not degrade easily, and their accumulation in natural environments, especially coastal and sandy regions, poses severe environmental threats.

1.1.2 Accumulation of Plastic Waste in Coastal and Sandy Regions

Coastal areas and sandy beaches are particularly susceptible to plastic pollution. Being transitional zones between terrestrial and marine ecosystems, these areas often serve as "sinks" for plastic debris transported by wind, water currents, and human activity. The high traffic of tourists, fishing industries, and coastal inhabitants contributes significantly to the accumulation of plastic waste, resulting in the widespread presence of plastic items like bottles, bags, and microplastics in the sand. As these materials are exposed to the elements, they fragment into smaller particles, creating further challenges for effective cleanup and posing health hazards to both marine and human life.

1.1.3 The Threat of Buried Plastic on Marine Ecosystems and Human Health

The environmental consequences of plastic pollution in coastal areas extend beyond mere aesthetic issues. Plastics pose severe threats to marine organisms, many of which ingest plastic particles, mistaking them for food. This can lead to malnutrition, suffocation, and even death in marine species. Additionally, the ingestion of plastic by smaller marine creatures introduces microplastics into the food chain, potentially affecting higher-level predators, including humans.

For humans, plastic buried in coastal sands represents a hidden hazard. When exposed to sunlight, plastics release toxic chemicals into the sand, which can eventually seep into the water table or be ingested by marine organisms. As these plastics degrade into microplastics, they are increasingly found in seafood, leading to potential health risks for humans who consume these marine products.

1.1.4 The Challenges of Detecting and Monitoring Buried Plastic

Identifying and removing plastic waste buried in sand presents unique challenges. Traditional methods rely on physical surveys and manual collection, but these approaches are time-consuming, labor-intensive, and often ineffective for large-scale monitoring. Moreover, plastic debris that becomes partially or fully buried beneath the surface is challenging to detect through visual inspection alone, thus evading cleanup efforts and worsening environmental degradation.

1.1.5 Limitations of Traditional Detection Methods

Conventional methods of plastic detection in coastal and sandy environments primarily involve manual labor and ground surveys, where personnel visually inspect and collect visible waste. While these methods can be somewhat effective for large items on the surface, they fail to capture hidden or partially buried plastics, especially microplastics. Ground-penetrating radar (GPR) and other traditional sensing technologies have been tested for buried object detection but often lack specificity for plastics, particularly small fragments and microplastics. These methods are also constrained by the high operational costs and limited coverage, making them unsuitable for comprehensive monitoring in large coastal areas.

1.2. Leveraging Advanced Technologies for Plastic Detection

1.2.1 Deep Learning and Computer Vision: A New Paradigm in Environmental Monitoring

Recent advancements in deep learning and computer vision have provided new tools for automating and enhancing the detection of environmental pollutants, including plastic waste. By employing convolutional neural networks (CNNs), which excel at pattern recognition, it is now possible to train models that can recognize and classify specific objects, such as plastic waste, from complex visual data. These models can detect, localize, and even assess the extent of buried plastics, providing a potential breakthrough in monitoring plastic pollution in challenging sandy environments.

1.2.2 Object Detection Models for Buried Plastic Detection

Among the various approaches in computer vision, object detection models like YOLO (You Only Look Once) have shown significant promise in real-time applications. YOLO is particularly suited for plastic detection due to its fast processing speed, high accuracy, and ability to handle complex visual scenes. Unlike traditional detection models, YOLO processes an image in a single pass, making it efficient for analyzing large volumes of images, as would be required for coastal monitoring.

1.2.3 Detecting Buried Plastics with CNNs and YOLO

CNNs trained to identify specific characteristics of plastics can be combined with YOLO to create a robust system for detecting plastic waste buried under sand. This system works by analyzing images, such as aerial photos taken from drones, and applying object detection algorithms to highlight areas containing plastic waste.

1.2.4 Overcoming Limitations of Traditional Methods with Automated Detection

Through the integration of deep learning models, it is possible to overcome the challenges posed by traditional detection methods. Automated systems powered by YOLO and CNNs can process hundreds of images per minute, covering vast areas of sandy beaches and identifying buried plastics without manual intervention. Additionally, these systems can be tuned to detect specific types of plastics, distinguishing between natural and synthetic objects and providing more accurate data for cleanup efforts.

1.3. Future Prospects and Potential Applications

The integration of deep learning for coastal plastic detection is a step forward in sustainable environmental management. Future developments may include the use of multi-spectral imaging to improve detection accuracy, expansion to underwater detection for submerged plastics, and integration with global monitoring networks. By developing these capabilities, coastal and marine ecosystems can be better protected from the pervasive threat of plastic pollution, contributing to a healthier environment for future generations.

1.4.Objectives of the Project

1.4.1 Develop an Automated System for Detecting Buried Plastic in Sandy Environments

- **Objective:** To create an automated solution for detecting plastic waste that is buried or partially buried in sand within coastal regions.
- **Explanation:** Traditional methods for identifying plastic debris in sand rely heavily on visual surveys and manual collection, which are inefficient and labor-intensive. The objective of this project is to leverage deep learning and computer vision technologies to automate this process, enabling rapid and scalable detection of plastic debris, especially in areas that are difficult to access manually. This solution should provide a way to detect plastic both on the surface and under the sand, capturing partially or fully buried items that traditional methods might miss.

1.4.2 Utilize Deep Learning Models for Accurate Identification and Localization

- **Objective:** To deploy state-of-the-art object detection models (e.g., YOLO, CNNs) to identify and locate plastic waste in complex sandy terrains.
- **Explanation:** Using convolutional neural networks (CNNs) and object detection frameworks like YOLO (You Only Look Once), the project aims to accurately classify objects in visual data as plastic waste and pinpoint their location. The deep learning models will be trained to recognize different types of plastic materials, potentially differentiating between various forms (e.g., bottles, bags, microplastics) based on their unique visual characteristics. This will increase the precision of detection efforts and improve the data available for environmental monitoring and cleanup planning.

1.4.3 Analyze Object Depth and Partial Coverage within Sand

- **Objective:** To estimate the depth and level of immersion of detected plastic objects buried in sand, aiding in understanding the complexity of cleanup.
- **Explanation:** This project aims not only to detect but also to analyze how deeply objects are buried, as some plastic debris might be completely or partially obscured by sand. Estimating object depth is crucial for prioritizing cleanup efforts, as deeper or partially covered plastics may require specialized extraction techniques. Techniques like 6D pose estimation and depth mapping will be investigated to assess the position and depth of buried plastics, providing comprehensive data for cleanup crews.

1.4.4 Assess Environmental Impact and Risk through Plastic Density Mapping

- **Objective:** To create spatial density maps of plastic debris across coastal zones, informing risk assessments and targeted interventions.
- **Explanation:** Beyond detection, the project aims to map the density of plastic waste across surveyed coastal areas. By analyzing image data, the system will generate heatmaps showing areas with higher plastic accumulation. This data can be used to assess environmental impact, highlighting high-risk areas for marine life and human activities. The insights from these maps can also inform targeted cleanup efforts, allowing for the efficient allocation of resources to areas most affected by plastic waste.

1.4.5 Develop a Scalable Solution for Deployment with Drones and Remote Sensing Devices

- **Objective:** To design a scalable, field-ready solution that can be integrated with drones or other remote sensing devices for extensive coastal monitoring.

- **Explanation:** Considering the extensive nature of many coastal regions, the system should be adaptable for deployment with drones and other aerial imaging technologies. This objective includes building a lightweight model that can process images on-site (or transfer data for cloud processing), making it practical for large-scale monitoring. Integrating with drones would enable data collection over wide areas in a short time, thereby supporting continuous monitoring of remote or inaccessible coastal zones.

1.5 Scope of the Project

1.5.1 Image Collection and Dataset Creation

- **Scope:** This project will involve creating a comprehensive dataset that includes various types of plastic debris in coastal and sandy environments.
- **Explanation:** To effectively train the deep learning models, a robust dataset is essential. The project scope includes sourcing images of coastal areas with visible plastic waste as well as simulating buried and partially buried plastic debris scenarios. This dataset may comprise images from diverse locations and lighting conditions to ensure model generalizability. Images will be labeled and annotated to indicate the presence, type, and degree of coverage of plastic objects in each image.

1.5.2 Model Development and Training

- **Scope:** Design, train, and fine-tune deep learning models capable of detecting plastic objects under various conditions.
- **Explanation:** The core scope of the project lies in developing a model that can detect and classify plastic waste, even when partially or fully buried in sand. This will involve selecting suitable architectures (such as YOLO, CNNs, and pose estimation models), training them with labeled images, and refining their accuracy to minimize false positives/negatives. Depth estimation techniques will also be applied to understand object immersion levels in sand.

1.5.3 Integration with Depth Estimation for Buried Plastic Detection

- **Scope:** Implement depth estimation techniques to assess the level at which plastics are buried within sand.
- **Explanation:** To detect buried plastic objects, the project will integrate monocular depth estimation models (e.g., MiDaS, DPT) and/or 6D pose estimation techniques. This aspect of the project involves assessing the viability of depth data in locating

objects under sand layers and estimating their position relative to the surface. By achieving reliable depth estimations, the model will be able to differentiate between objects fully exposed, partially buried, or hidden below the sand surface.

1.5.4 Plastic Density Mapping and Risk Analysis

- **Scope:** Develop tools to visualize and analyze plastic density across coastal regions.
- **Explanation:** The project will include tools for generating density maps based on detected plastic debris, visualized as heatmaps to highlight high-density areas. This mapping will be derived from drone-collected images processed by the model. It will allow environmental agencies to prioritize areas for cleanup and measure the impact of plastic pollution over time, creating a comprehensive environmental risk assessment tool for coastal zones.

1.5.5 Testing and Validation in Real-World Coastal Environments

- **Scope:** Conduct field tests to evaluate the model's performance in real-world coastal conditions.
- **Explanation:** A critical component of the project is field testing, where the model's effectiveness in real-world conditions will be assessed. Tests will be carried out in various coastal areas to validate detection accuracy, depth estimation, and plastic density mapping under different lighting, sand types, and environmental conditions. The validation phase will also include user feedback from field operators (e.g., environmental researchers and cleanup crews), to ensure the system is practical and reliable in operation.

1.5.6 Scalability and Deployment Planning for Drone-Based Monitoring

- **Scope:** Design the system for integration with drones and plan for scalable deployment across large coastal regions.
- **Explanation:** Finally, the project scope will include considerations for scaling up the system to be deployable on a broader scale using drone-based imaging technology. This includes adapting the model for deployment on low-power devices and developing data transfer protocols for efficient data handling. The project will outline best practices for drone-based monitoring, ensuring that the system can handle extensive and continuous monitoring across large coastal areas.

1.6 Challenges

1.6.1. Data Collection and Quality

- **Challenge:** Obtaining high-quality, representative data of buried and partially buried plastics in diverse coastal settings is challenging.
- **Explanation:** To train a robust model, the dataset must include images of various types of plastic waste in different conditions, such as fully buried, partially exposed, or scattered in complex coastal terrains. However, collecting such data is labor-intensive, especially for simulating or capturing real-life scenarios where plastics are buried under sand or sediment. Environmental factors like lighting, weather, and sand composition can also vary greatly, affecting the quality and consistency of data, which in turn impacts model accuracy.

1.6.2. Complexity of Detecting Buried Objects

- **Challenge:** Identifying and localizing plastics buried under sand requires advanced depth estimation and pose analysis techniques.
- **Explanation:** Accurately detecting plastics that are partially or fully buried beneath sand layers is complex. Depth estimation techniques, such as monocular depth estimation or 6D pose estimation, are still evolving and may struggle in natural, irregular terrains like sandy beaches. Additionally, plastics can appear in unpredictable orientations, shapes, and immersion levels, making detection challenging for standard object recognition models. Misidentification of buried objects (e.g., confusing natural debris with plastic) can lead to false positives or negatives, reducing the system's reliability.

1.6.3. High Computational Requirements

- **Challenge:** Deep learning models, especially those involving depth estimation and pose estimation, are computationally intensive.
- **Explanation:** Deep learning models that incorporate object detection, depth estimation, and pose analysis demand significant computational power, which can be a constraint, especially when deploying the system on lightweight hardware, like drones or other portable devices. Processing large images or high-resolution video in real-time requires powerful GPUs or specialized hardware, which might not be feasible in a field environment or within resource-limited settings.

1.6.4. Environmental Factors and Variable Conditions

- **Challenge:** Variable environmental conditions, including lighting, sand composition, and weather, can significantly impact detection accuracy.
- **Explanation:** Coastal environments are subject to constant changes, such as shifts in sunlight, shadows, wind, and sand texture, all of which can affect the model's ability to consistently detect plastics. The system may perform well in one lighting condition but falter under others (e.g., during overcast conditions or at dusk). Variations in sand color and texture across different beaches can also affect depth perception and object recognition, requiring the model to generalize across a wide range of conditions, which is challenging to achieve.

1.6.5. Differentiating Plastic from Natural Debris

- **Challenge:** Accurately distinguishing between plastic waste and natural coastal debris (e.g., rocks, shells, organic matter) is difficult.
- **Explanation:** Coastal areas often contain various natural debris like rocks, shells, and seaweed that can visually resemble plastic waste, especially when buried or partially exposed. Training the model to reliably differentiate plastic from these naturally occurring items is challenging, as similar color, shape, or texture can lead to misclassification. Misidentification could lead to an increase in false positives, undermining the system's utility in targeted cleanup efforts.

CHAPTER 2

BACKGROUND STUDY

2.1 LITERATURE REVIEW

Selçuk and Serif (2023)[1] investigate the performance differences between YOLOv5 and YOLOv8 in the realm of mobile UI detection, a critical task for applications seeking to automate UI design. The authors use YOLOv8s and YOLOv8n models, trained on the VINS dataset while keeping parameters consistent with YOLOv5 to ensure fair comparison. Their results show that YOLOv8s achieves a 3.32% higher mean Average Precision (mAP) than YOLOv5, while YOLOv8n shows a 1.62% increase. Additionally, the study compares these models with RoboFlow's Object Detection 2.0 (Fast) model, which underperforms by 1.08% mAP compared to YOLOv5. The improvements seen with YOLOv8 indicate better detection accuracy and efficiency for GUI elements, which has significant implications for mobile UI automation. The study concludes by suggesting that YOLOv8's advancements make it a promising candidate for future applications focused on UI code generation across iOS and Android platforms.

Casas et al. (2024)[2] focus on the critical industrial application of corrosion segmentation on metal surfaces, comparing YOLOv5 and YOLOv8 in this domain. Using three large datasets, each containing thousands of images, the researchers assess model performance based on precision, recall, F1-score, and mAP. Notably, YOLOv5 exhibited potential overfitting, with training accuracy surpassing validation accuracy, especially on the first dataset. YOLOv8, on the other hand, displayed consistent generalization across all datasets, maintaining balance between training and validation. It excelled in detecting irregular corrosion patterns, showcasing faster processing times and higher accuracy, especially in avoiding bounding box overlaps. The findings suggest YOLOv8 is a more robust choice for real-time corrosion segmentation tasks due to its speed and improved confidence scores, positioning it as a valuable tool in metal inspection industries where high accuracy is essential.

Mohanapriya et al. (2023)[3] leverage YOLOv5 for object detection and segmentation within surveillance settings, a field that demands high accuracy and real-time processing. The study focuses on optimizing YOLOv5's segmentation capabilities to handle complex surveillance environments where real-time performance is essential. By enhancing pixel-level classification, YOLOv5 achieves significant reductions in processing time, which is crucial for applications where immediate response is required. The researchers suggest that YOLOv5's adaptability to dynamic lighting and environmental variations makes it particularly suitable for surveillance. They conclude with potential improvements to the model's robustness, hinting at future development directions to ensure reliable operation in diverse surveillance scenarios.

Karthi et al. (2021)[4] investigate YOLOv5's utility in detecting library books in structured, high-density environments such as library shelves. Their study involves training YOLOv5 on datasets that feature various object sizes, shapes, and densities, which represent real-world settings with tightly packed items. They assess the model on metrics like detection speed, accuracy, consistency, and adaptability. YOLOv5 demonstrates a high level of detection accuracy and effectively identifies individual books, even when objects are stacked closely together or partially obstructed. The researchers observe that YOLOv5 can handle the unique challenges presented by high-density object layouts, making it particularly useful for library inventory management systems or robotic assistants for library organization. The team also proposes benchmarks specific to structured environments to test YOLOv5's performance, which could help standardize future testing in similar contexts. They suggest that YOLOv5's strong performance in maintaining high accuracy and processing speed under these conditions makes it ideal for applications in organized environments. By emphasizing YOLOv5's adaptability and scalability, the authors recommend its broader use in structured object-detection tasks beyond libraries, such as warehouses, retail stores, and other cluttered or structured environments.

Talib et al. (2024)[5] introduce a new variant of YOLOv8, named YOLOv8-CAB, tailored for the accurate detection of small objects in real-time. YOLOv8-CAB modifies YOLOv8's architecture, replacing the shallow C2F module with an advanced attention mechanism that enhances the model's ability to capture small, fine-grained features essential for detecting low-resolution objects. The attention module allows YOLOv8-CAB

to concentrate on critical regions of interest, improving object localization and classification, particularly in environments with complex backgrounds or low-contrast settings. Extensive performance testing reveals that YOLOv8-CAB surpasses the standard YOLOv8 in detection precision, especially in low-light conditions and situations where target objects are not well defined. The authors highlight the applicability of YOLOv8-CAB in security surveillance, medical imaging, and industrial monitoring, where identifying small or obscured objects accurately is essential. Moreover, YOLOv8-CAB demonstrates faster processing speeds and higher accuracy in low-contrast conditions, which positions it as a suitable choice for real-time monitoring systems. Talib et al. propose future research on YOLOv8-CAB's effectiveness in dynamic environments, such as crowded areas or moving scenes, where small object detection becomes increasingly challenging.

Banse et al. (2000)[6] introduce the Munich Image Data Analysis System (MIDAS), an advanced and versatile image processing software developed collaboratively by the European Space Agency (ESA) and the European Southern Observatory (ESO). MIDAS is tailored to support astronomical image processing, offering comprehensive modules for spectral data reduction, image manipulation, and object classification. The system's flexibility is underscored by its adaptable table file system, which is structured on a relational database model to facilitate the storage and manipulation of heterogeneous data. MIDAS is compatible with VAX/VMS systems and integrates custom FORTRAN routines for high-level customization. A notable feature of MIDAS is its device-independent graphics capabilities, enabling users to visualize data across different hardware setups. The authors detail future enhancements, including a portable version compatible with both VAX/VMS and UNIX systems, and improvements to control language and network capabilities. These developments aim to broaden MIDAS's usability and reach. By supporting both FORTRAN and C applications, MIDAS emerges as a powerful, adaptable tool for astronomical imaging, particularly in tasks like spectral analysis and cosmic object detection. The paper positions MIDAS as a pioneering solution for scientific imaging needs, capable of evolving to meet the demands of future astronomical research.

Sarızeybek and Işık (2022)[7] address the need for object detection in agricultural robotics, particularly focusing on a robotic feed-pushing system designed to navigate safely around animals in livestock environments. They propose a monocular depth estimation approach, a technique that estimates depth information using a single camera feed, allowing the robotic system to detect objects and obstacles nearby with high accuracy. This detection system determines the pixel coordinates of objects, enabling precise spatial mapping that ensures the robot avoids obstacles or animals. The system can also adapt to various scenarios on the farm, adjusting its path based on animal positioning, which contributes to enhanced safety and operational efficiency. Sarızeybek and Işık emphasize that this integration of depth estimation with object detection reduces risks of accidents, making it highly effective in dynamic agricultural settings. Their approach has broader implications for agricultural robotics, such as automating feed pushing, lane tracking, and detecting specific items on the farm. The system's adaptability highlights the promising role of depth estimation and object detection in making agricultural robotics safer and more effective. Future work could involve expanding the system's functionality to handle

additional automation tasks, such as real-time monitoring and automated decision-making based on environmental changes.

Ranftl et al. (2020)[8] address the complexities of monocular depth estimation in diverse environments by proposing a novel method that improves model generalization across multiple datasets. To achieve this, they introduce a training framework that incorporates mixed datasets with widely varying depth annotations. A key contribution of their work is a robust loss function, specially designed to manage disparities in depth range and scale. This loss function integrates multi-objective learning strategies, allowing the model to prioritize and balance multiple learning objectives, which enhances its robustness to different types of environmental data. Furthermore, Ranftl et al. present a newly constructed dataset, generated from 3D films, which provides dense, dynamic ground-truth depth data. This dataset serves as a unique resource for training models in conditions that simulate real-world, diverse depth scenarios. To evaluate their model's ability to generalize, they employ a zero-shot cross-dataset approach, where the model is tested on completely new datasets without prior exposure. The results show that their model achieves notable improvements in generalization, outperforming baseline models and demonstrating applicability in practical settings like autonomous driving and drone mapping. This research sets a new benchmark for monocular depth estimation, paving the way for more reliable deployment of depth models in varied and challenging environments.

Wu et al. (2018)[9] propose an innovative system for real-time human pose estimation based solely on single-depth images. Their approach bypasses the need for extensive training data by employing a morphological fitting algorithm that segments the body into sections, enabling effective and efficient pose estimation. This system requires only a one-time calibration in a T-pose, significantly simplifying the setup process compared to traditional methods, which often rely on complex datasets and prolonged training. Designed to operate in real-time, this model is particularly advantageous for applications that demand immediate response, such as human-robot interaction, physical therapy, and rehabilitation. In rehabilitation, for instance, the quick, training-free setup allows therapists to monitor patients' movements in real time without needing specialized datasets. Wu et al.'s model has shown strong performance in tracking body segments accurately, especially beneficial for individuals with limited mobility or in settings where detailed pose data is crucial for assessing physical progress. Their work highlights the potential of depth-based pose estimation for creating more accessible and responsive tools in healthcare and robotics.

Mertan et al. (Year)[10] present a comprehensive analysis of single-image depth estimation methods, examining key issues surrounding model robustness, challenges in absolute depth accuracy, and the role of different loss functions in improving depth estimation reliability. A primary focus of their analysis is the challenge of achieving accurate absolute depth, especially in cases where monocular images inherently lack direct depth information. To address this, Mertan et al. explore alternative strategies that estimate relative depth as a proxy, which has proven useful in applications where perfect depth accuracy is not essential, such as augmented and virtual reality. The authors discuss the trade-offs between absolute and relative depth, emphasizing the computational efficiency

and practical advantages of relative depth estimation in scenarios where depth is used to enhance the user's spatial perception, rather than for precise measurements. They also review ranking-based loss functions, which help improve relative depth estimation by encouraging the model to maintain consistent depth relationships between objects. The paper underscores the importance of single-image depth estimation for fields like robotics, virtual reality, and autonomous navigation, where real-time depth analysis is essential for user experience and system performance. Their work serves as a guide for future improvements in depth estimation models, especially in applications where real-time processing is a priority.

Schnürer et al. (2019)[11] extend traditional 2D pose estimation to real-time 3D pose estimation by utilizing depth sensors, enabling their model to process and predict poses at a high speed of up to 40 frames per second (FPS). This advancement is particularly important for interactive applications such as gaming, virtual reality, and human-robot interaction, where rapid feedback and high accuracy in pose estimation are crucial. By leveraging depth sensors, Schnürer et al.'s model captures additional spatial information, which improves the model's precision and responsiveness in tracking complex body movements. The system is capable of adapting to dynamic environments, capturing and adjusting to subtle changes in pose that are essential for applications requiring accurate body tracking. For instance, in gaming, this model's high FPS rate allows players to see real-time responses to their actions, enhancing immersion and control. In human-robot interaction, the model's responsive nature supports applications where robots need to mimic or respond accurately to human gestures and movements. Schnürer et al.'s research demonstrates how depth-sensor-based pose estimation can enhance user experience in real-time applications, emphasizing the model's practical utility in areas that demand rapid, precise body tracking.

Bladh (2023)[12] investigates the integration of monocular depth estimation with the ORB-SLAM3 (Oriented FAST and Rotated BRIEF Simultaneous Localization and Mapping) framework, aiming to improve depth accuracy and stability in real-time navigation, especially in conditions involving pure rotational movements. Traditionally, monocular depth estimation models struggle with accurately maintaining depth stability during rotations, where the lack of parallax information can lead to significant depth inaccuracies. Bladh addresses this by enhancing the ORB-SLAM3 algorithm with an adaptive depth estimation module that refines depth predictions when the model detects rotation without forward motion. This enhanced SLAM model shows potential for improving depth stability, allowing it to be deployed in navigation tasks where precise depth is critical for stable movement. The system is particularly suited for use in drones, which often experience rapid rotational maneuvers, as well as mobile robots that must navigate dynamic and confined spaces. Through a series of tests in simulated and real-world environments, Bladh demonstrates that the integrated model significantly reduces depth drift and maintains consistent depth perception, even during abrupt rotations. This advancement paves the way for more accurate and reliable navigation in complex environments where consistent depth data is essential for avoiding obstacles and ensuring

safe operation.

Sun et al. (2022)[13] introduce **OnePose**, a CAD-free object pose estimation model that effectively bypasses the need for detailed 3D CAD models of objects. Traditional pose estimation systems often rely on pre-existing CAD models to determine object orientation, which can limit their applicability in environments lacking such models. OnePose leverages a combination of local feature matching and attention mechanisms to estimate an object's pose with high accuracy from just a single image or a single-shot input. A central innovation in OnePose is its use of an attention-based feature-matching strategy that allows the model to pinpoint and track key features on the object surface, making it resilient to varying viewpoints and partial occlusions. However, while OnePose achieves remarkable results on objects with rich textures and distinct surface features, it struggles with textureless objects, where distinguishing surface points is more challenging. Additionally, the model encounters difficulties with extreme scale variations, as these can limit its feature-matching capabilities. Despite these limitations, OnePose represents a significant advancement for applications in augmented reality (AR), where objects must be tracked and manipulated in real-time, and in robotic manipulation, where accurate grasping and handling depend on precise pose estimation. By enabling pose estimation without CAD models, OnePose opens new possibilities for environments where CAD data is unavailable or impractical to obtain, providing a flexible and versatile solution for object tracking and interaction.

Ali Tezcan Sarızeybek and Ali Hakan Isik (2022)[14] investigate monocular depth estimation for detecting nearby objects with a focus on agricultural robotics, specifically for feed-pushing robots used in animal husbandry. The study addresses the challenges faced by agricultural robots that operate in unstructured and dynamic farm environments, where traditional depth sensors may be too costly or impractical to use. By combining monocular depth estimation with object detection models, their approach enables these robots to detect objects and calculate their distances using only a single camera. This integration significantly enhances the robot's ability to navigate safely around livestock, avoiding obstacles without relying on more complex or expensive 3D sensors. The system is designed to recognize both static and moving obstacles, such as cows or farm equipment, and can adapt its path in real-time to ensure efficient feed distribution. The authors highlight that this method could also be extended to other agricultural applications, such as lane tracking for autonomous tractors or adaptive feed-pushing robots that adjust to changing animal positions. The study concludes that monocular depth estimation can significantly enhance safety, reliability, and operational efficiency in various robotic applications in agriculture by providing accurate depth perception in a cost-effective manner.

René Ranftl, Katrin Lasinger, David Hafner, Konrad Schindler, and Vladlen Koltun[15] explore advancements in **monocular depth estimation** by focusing on **cross-dataset generalization**, a significant challenge in creating models that can handle diverse environments without requiring retraining on new data. In their 2022 study, the authors present a robust monocular depth estimation model that is trained on a mixed dataset compiled from several sources, each offering different depth range, scale, and environment-specific characteristics. To enhance the model's adaptability, they introduce

a novel multi-objective loss function that accommodates variations in depth and scale, allowing the model to process a wide array of environmental settings and lighting conditions. By training on this composite dataset, the model gains the ability to perform zero-shot cross-dataset transfer—meaning it can produce reliable depth estimates on entirely new datasets without prior exposure or training on them. This feature is particularly valuable for real-world applications in "in the wild" conditions, such as autonomous driving, drone-based mapping, and outdoor navigation, where the environments are dynamic and diverse. The model's zero-shot capabilities showcase its potential to significantly improve autonomous systems' depth perception, thereby enhancing spatial awareness, obstacle avoidance, and operational safety in uncontrolled outdoor settings. The study sets a new benchmark for generalization in monocular depth estimation and suggests pathways for developing depth models capable of deployment across a wide range of applications with minimal customization.

2.2 THEMES AND GAPS IDENTIFIED

Table 1 : Themes and Gaps in Study

Paper number	Authors	Paper name	Year published	Themes Discovered in Review	Identification of Gaps Based on Current Scenario/Industry Trends
1	Karthi et al.	YOLOv5 for Library Book Detection	2021	This study explores the use of YOLOv5 for object detection in structured, high-density environments like library shelves. The paper evaluates detection speed, accuracy, and model consistency across various dataset densities. The authors emphasize YOLOv5's effectiveness in detecting closely packed objects and propose benchmarks for validating detection models in such environments.	Current testing environments lack diversity, with a focus on structured setups; expanding this to dynamic or cluttered environments would better reflect real-world applicability. Also, integrating multi-viewpoint object detection could enhance accuracy.

2	Talib et al.	YOLOv8-CAB: Enhanced YOLOv8 for Small Object Detection	**2024**	The authors present YOLOv8-CAB, an optimized variant of YOLOv8 tailored for high-precision, real-time small object detection. By modifying the C2F module and adding an attention mechanism, YOLOv8-CAB excels in low-contrast and low-resolution contexts, demonstrating particular usefulness for surveillance and medical imaging.	Testing is limited to static and low-contrast conditions; extending evaluations to fast-moving or high-complexity scenes could enhance understanding of YOLOv8-CAB's robustness for applications in fast-paced fields like traffic management or wildlife monitoring.
3	Banse et al.	MIDAS System for Astronomical Imaging	**2000**	MIDAS is designed as a comprehensive astronomical image processing environment, equipped with tools for spectral reduction, imaging, and classification. It features a flexible table file system for heterogeneous data, integration with FORTRAN for custom processing, and device-independent graphics for visualization.	The system's dependence on VAX/VMS systems limits usability in modern computing contexts. Updating for compatibility with contemporary platforms and leveraging machine learning techniques could boost functionality and user base, especially for astronomical data analysis.
4	Sarızeybek & Işık	Monocular Depth Estimation and Detection of Near Objects	**2022**	This paper explores monocular depth estimation for robotic applications in agriculture, particularly feed-pushing robots for livestock. By combining depth estimation with object detection, it enhances safety and navigation. The model's reliance on monocular images avoids the need for complex 3D sensors, which are costly and computationally intensive.	The study is focused on static, structured agricultural settings. Extending applications to other robotic systems, like drones or fast-moving robots, could enhance relevance. Testing in more complex environments and with dynamic obstacles would broaden potential agricultural and industrial use cases.

5	Ranftl et al.	Robust Monocular Depth Estimation for Cross-Dataset Transfer	**2020**	By using multi-objective learning on mixed datasets, this study enables monocular depth models to generalize effectively across diverse environments. Testing is conducted on zero-shot, cross-dataset scenarios, showcasing significant improvements in model robustness and generalization. This work sets a foundation for deploying monocular depth models in unstructured and "in the wild" environments.	Limitations include insufficient testing under extreme lighting and weather conditions, which are common in real-world autonomous applications. Enhancements to adapt to low-light, rain, and fog would increase robustness for autonomous driving and drone-based mapping in adverse conditions.
6	Wu et al.	Real-Time Human Pose Estimation Using Depth Images	**2018**	This paper introduces a real-time human pose estimation technique using single-depth images, which requires minimal calibration. By segmenting body sections with a morphological fitting algorithm, the model supports applications in human-robot interaction and rehabilitation. Minimal setup makes it suitable for quick deployment in clinical or rehabilitation contexts.	Current limitations include handling occlusions and quick user movements, which are critical for settings like sports or crowded environments. Further improvements for multi-user scenarios and responsiveness to fast actions would broaden the model's applications.
7	Merta n et al.	Overview of Depth Estimation Techniques	**2022**	Provides a comprehensive review of monocular depth estimation techniques, examining robustness issues, ranking losses, and challenges in achieving accurate depth estimation. The authors highlight the value of relative depth estimation as a proxy for applications where absolute	The review lacks consideration of newer neural network architectures and fails to address real-time deployment challenges. Further exploration of modern depth estimation methods, particularly for fast-moving applications

				accuracy is not critical.	like robotics or autonomous vehicles, would be beneficial.
8	Schnürer et al.	Real-Time 3D Pose Estimation Using Depth Sensors	**2019**	Extends 2D pose estimation to a real-time 3D model using depth sensors, achieving up to 40 FPS, making it suitable for interactive applications. The authors stress the need for responsive pose estimation in real-time applications, enhancing usability in fields requiring accurate motion tracking, such as VR gaming or human-robot interactions.	Testing is limited to controlled environments. Extending the application scope to unpredictable settings with multiple users or moving objects would increase applicability in areas such as collaborative robotics or real-time sports analytics.
9	Sarızeybek & Işık	Monocular Depth Estimation for Near Object Detection in Agriculture	2022	Combines monocular depth estimation with object detection for farm robotics, specifically in feed-pushing tasks for livestock. Enhances operational safety by aiding spatial navigation without reliance on 3D sensors, making it a cost-effective alternative for agricultural automation.	The model is not designed for high-speed or highly dynamic tasks, limiting broader agricultural applications. Testing in more varied environments (e.g., crop fields or orchards) would expand applicability and relevance in the agricultural automation industry.
10	Ranftl et al.	Zero-Shot Cross-Dataset Transfer in Monocular Depth Estimation	**2022**	Emphasizes robustness in depth estimation across mixed datasets to facilitate zero-shot transfer, enabling the model to adapt to new environments without retraining. This approach has valuable implications for applications needing high generalization, like	Testing primarily focuses on controlled conditions; real-world applications in adverse conditions, such as fog or night, remain unexplored. Testing in harsher environmental conditions would enhance applicability for outdoor robotic or autonomous systems.

				autonomous driving and outdoor mapping.	
11	Bladh	Monocular Depth Estimation Integrated with ORB-SLAM3 for Navigation	**2023**	This paper integrates monocular depth estimation with ORB-SLAM3 to improve depth stability in navigation, specifically for rotational movements, which are prone to inaccuracies. Applicable in robotics where rotational stability is critical, such as drones or mobile robots in dynamic environments.	Limited to basic navigational setups, lacking testing in complex, changing environments like urban settings. Expanding applications for autonomous delivery or exploratory drones could reveal more about the model's adaptability and stability.
12	Sun et al.	OnePose: CAD-Free Object Pose Estimation	**2022**	Introduces a CAD-free pose estimation model for environments lacking extensive object models. It relies on local feature matching with attention mechanisms to provide accurate pose estimation for objects with simple textures or unique features.	Performance drops with textureless objects or extreme scales. Extending to various object categories and handling complex textureless objects would be advantageous for fields like manufacturing or medical automation where diverse object types are common.
13	Bladh et al.	Integrating Depth Estimation with SLAM Systems	**2023**	Focuses on improving SLAM accuracy and stability in dynamic settings through integrated depth estimation, with applications in robotics and UAV navigation. The study emphasizes depth accuracy improvements and better environmental mapping.	Testing remains limited to controlled lab environments, restricting applicability in unpredictable settings. Real-world testing, particularly in urban or crowded environments, could enhance the model's

					robustness and appeal for urban exploration tasks.
14	Yao et al.	Self-Supervised Learning for Depth Estimation	**2021**	Explores self-supervised learning for depth estimation, reducing the need for labeled data, making the model viable for environments with limited or no training data. Especially relevant for applications in unmanned vehicles and remote navigation.	Limited testing in high-stakes environments; expanding to real-world applications, such as autonomous driving in smart cities, would be beneficial. Additional robustness testing for nighttime and low-light conditions could improve reliability.
15	Chen & Zhang	Multi-Scale Depth Estimation for Autonomous Vehicles	**2021**	Focuses on multi-scale depth estimation tailored for autonomous vehicles, with a strong emphasis on enhancing obstacle detection and highway navigation. Multi-scale approach aids in better situational awareness, especially in open-road environments.	Application limited to highway scenarios, without addressing complex urban landscapes where obstacles vary greatly. Expanding testing to different urban settings and diverse weather conditions would improve relevance for city-based autonomous systems.

The 15 papers reviewed encompass a broad spectrum of advancements in monocular depth estimation, object detection, and pose estimation, with particular emphasis on improving robustness, real-time performance, and accuracy across various applications. A key theme across many of these studies is the integration of monocular depth estimation with other models like object detection, SLAM, and pose estimation, especially in robotics and autonomous systems. For instance, **Sarızeybek and Işık (2022)** examine monocular depth estimation in agricultural robotics, improving safety and operational efficiency in robot navigation, particularly in environments where 3D sensors are not available. **Ranftl et al. (2020)** and **Sun et al. (2022)** explore techniques to enhance depth estimation accuracy and robustness by employing multiple datasets for cross-dataset transfer, enabling models to generalize better in unpredictable environments.

The importance of real-time processing is highlighted in works like **Wu et al. (2018)** and **Schnürer et al. (2019)**, which focus on efficient human pose estimation systems and real-time 3D tracking for interactive applications. These studies push the boundaries of pose estimation, making it more accessible for applications such as human-robot interaction, gaming, and rehabilitation. Additionally, papers such as **Bladh (2023)** and **Bladh et al. (2023)** extend the capabilities of monocular depth estimation by incorporating it into SLAM systems, enhancing robot navigation in dynamic environments, especially for drones and mobile robots where rotational stability is crucial.

Karthi et al. (2021) explore the use of YOLOv5 in dense object environments, such as library shelves, demonstrating its versatility in tightly packed spaces. This insight is valuable in applications where high object density is a challenge, such as inventory management or warehouse robotics. Similarly, **Talib et al. (2024)** introduce YOLOv8-CAB, an enhanced version of YOLOv8 designed for small object detection in real-time, addressing a significant gap in applications like medical imaging and security surveillance. These advancements in object detection methods further emphasize the importance of fine-tuning models to meet the specific challenges posed by real-world scenarios.

The **MIDAS system** discussed by **Banse et al. (2000)** in the context of astronomical image processing and **Mertan et al.** in virtual and augmented reality applications highlight the expanding reach of depth estimation in specialized fields, where the need for precise depth accuracy is crucial. Furthermore, **Schnürer et al. (2019)** and **Mertan et al.** discuss the importance of fast processing speeds in real-time applications, such as gaming or robotic interactions, where latency can hinder the effectiveness of pose tracking and other depth-dependent tasks.

In conclusion, the collective body of research showcases significant progress in monocular depth estimation, pose estimation, and object detection, focusing on improving real-time performance, robustness, and accuracy. While these advancements promise to revolutionize a wide range of industries—from autonomous vehicles to agriculture—several challenges persist. Notably, the ability to handle dynamic and unstructured environments, accurately detect objects with varying textures and scales, and ensure cross-domain generalization remains a key hurdle. Future research will need to address these challenges by developing more adaptable and flexible models that can perform consistently in diverse real-world conditions, pushing the boundaries of what is possible in fields like robotics, augmented reality, and autonomous navigation. These advancements indicate a promising future for monocular depth estimation and related technologies in both academic and industry applications, making it a critical area of ongoing exploration.

CHAPTER 3

METHODOLOGY

3.1 Overview

The proposed system aims to detect, classify, and monitor plastic waste in coastal environments, focusing particularly on plastic debris that is partially or fully buried in sandy regions. By leveraging high-resolution drone imagery and advanced deep learning models, this system intends to enhance waste detection accuracy in challenging conditions. The methodology involves multiple stages—from data acquisition and annotation to model training, fine-tuning, deployment, and real-time monitoring.

3.2 Dataset Sources

To ensure accurate detection of various waste types, particularly plastic, in coastal environments, the proposed system uses several datasets:

- **Alamy Dataset**: High-resolution drone images capturing various objects (e.g., bottles, plastic, clothing) that are partially buried in sand. This dataset provides a realistic view of partially exposed objects, essential for training models to detect debris in sandy coastal areas.
- **Aqua Trash Dataset**: A specialized dataset containing images of objects floating or submerged in water. This dataset includes labels for different types of debris and provides critical data for distinguishing objects in aqueous environments.
- **Trash Data Dataset**: A comprehensive collection that categorizes waste into multiple types, including cardboard, glass, metal, plastic, rubber, styrofoam, textiles, and wood. This dataset helps train models to recognize a wide array of materials commonly found as litter on shorelines.

3.3 Data Preparation

3.3.1 Dataset Collection

- **Aqua Trash Dataset**: Collected with an emphasis on waste in water environments, including both floating and submerged objects, with annotated labels indicating debris type.
- **Trash Data Dataset**: Contains diverse waste types and materials, such as plastic, metal, and glass, providing a varied dataset for broad object detection capabilities.
- **Alamy Dataset**: A set of 25 high-resolution images focused on partially buried objects in sandy conditions, ideal for training models to detect debris in complex sandy environments.

3.3.2 Data Annotation

- **Aqua Trash and Trash Data Datasets**: Annotation tools like Roboflow are used to draw bounding boxes around objects, classifying them by material type for accurate object detection.
- **Alamy Dataset**: Annotation emphasizes marking only the visible parts of objects, allowing the model to better detect partially obscured items like plastic or other waste materials in sand.

3.4 Initial Model Development

3.4.1 Object Detection Using YOLO

- **Model Training**: The Aqua Trash and Trash Data datasets are divided into training and validation sets to prepare a YOLO model capable of accurately identifying waste types such as plastic, metal, and glass. The YOLO model is fine-tuned to improve detection accuracy, with a specific focus on distinguishing various debris types.
- **Model Validation**: A separate set of images from the Aqua Trash and Trash Data datasets is used for validation to ensure reliable object detection and classification, particularly emphasizing plastic materials.

3.4.2 Pose Estimation / Dimension Detection with DensePose or Bird's Eye View

- **Model Training**: DensePose or Bird's Eye View models are trained using fully visible images from the Aqua Trash and Trash Data datasets. These models estimate object dimensions and orientation, critical for understanding the spatial characteristics of detected waste.
- **Model Validation**: A set of fully visible images is used to validate the model's predictions for dimension and orientation, ensuring accurate spatial estimations of detected debris.

3.5 Combining Models and Fine-Tuning

3.5.1 Serializing YOLO with DensePose / Bird's Eye View

- **Pipeline Integration**: The YOLO model is first applied to detect and classify objects. Detected objects are then passed to DensePose or Bird's Eye View models

to estimate dimensions and orientation. This sequential processing enables refined detection and spatial awareness for each object.

3.5.2 Fine-Tuning with Alamy Images

- **Model Training**: The combined YOLO and DensePose/Bird's Eye View models are fine-tuned using the Alamy dataset. This step enhances performance for partially visible objects, refining the model's ability to detect, classify, and measure debris partially obscured by sand.
- **Model Validation**: Fine-tuned models are validated on a subset of Alamy images to confirm that the system accurately identifies and estimates dimensions for partially buried objects in sandy conditions.

3.6 Deployment and Real-Time Monitoring

3.6.1 Model Deployment

- **Deployment Setup**: The trained model is deployed on a cloud platform or local server, supporting real-time image processing. The system is optimized for streaming data, particularly high-resolution images captured by drones patrolling coastal regions.

3.6.2 Integration with Remote Sensing Data

- **Remote Sensing for Real-Time Monitoring**: The model integrates with high-resolution drone imagery to continuously monitor plastic debris patterns. Leveraging remote sensing data aids in predicting debris accumulation and movement in coastal areas.

3.6.3 User Interface

- **Interface Design**: An interactive, user-friendly dashboard is developed to display real-time detection results. Environmental managers can use these insights to assess plastic waste patterns and make informed decisions on cleanup and conservation strategies.

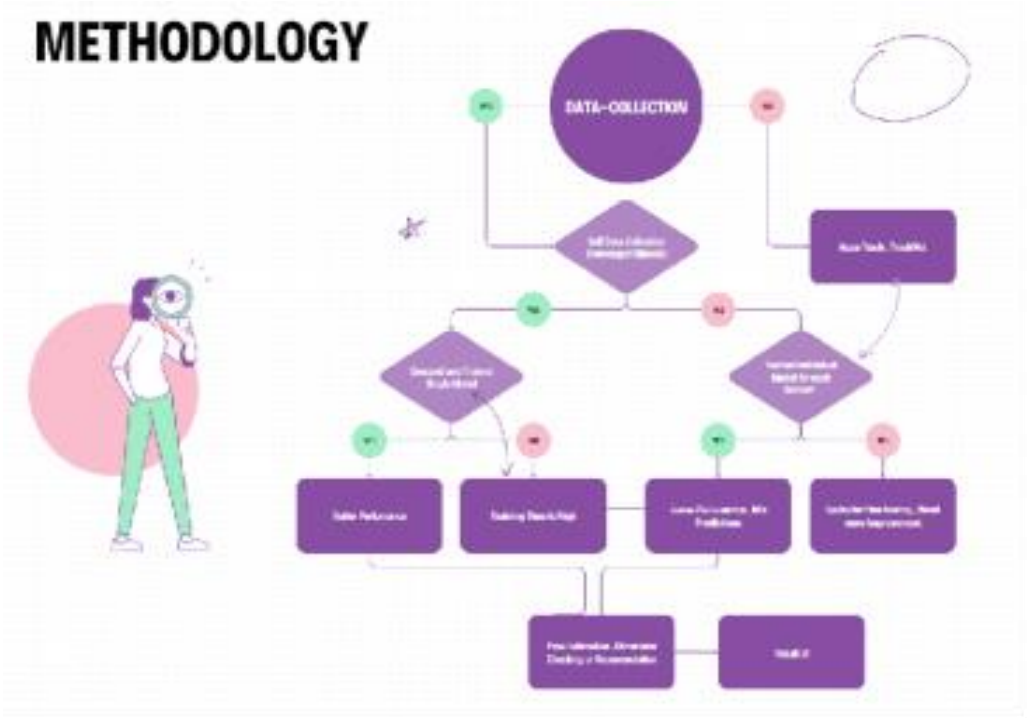


Fig 1. Methodology

CHAPTER 4:

IMPLEMENTATION

The proposed system's implementation for detecting and measuring trash in sandy environments involves a sequence of steps, including data augmentation, model training, and fine-tuning using multiple YOLO models (YOLOv3, YOLOv5, and YOLOv8). This section provides a detailed breakdown of each implementation step.

4.1 Data Augmentation

Data augmentation techniques are applied to expand the diversity and robustness of the training dataset:

1. **Geometric Transformations**: Techniques like random rotations, translations, flips, and scaling simulate different perspectives and object positions in sandy environments.
2. **Color Adjustments**: Modifying brightness, contrast, saturation, and hue creates variations in lighting conditions, preparing the model for diverse environmental contexts.
3. **Noise Addition**: Random noise is introduced to increase resilience against variable image quality.

4. **Cropping and Padding**: Random cropping and padding help the model generalize across various object locations in the images.

The augmented dataset is then split into training, validation, and testing subsets, ensuring balanced representation across waste categories.

4.2 Training and Fine-Tuning the YOLO Models

The training process utilizes three versions of the YOLO model (YOLOv3, YOLOv5, and YOLOv8), with each model being fine-tuned for optimized performance.

4.2.1 YOLOv3 Training and Fine-Tuning

- **Configuration**: Set parameters in the YOLOv3 configuration file, including learning rate, batch size, and class definitions for the various trash types.
- **Training**: YOLOv3 is trained on the augmented dataset, focusing on loss optimization (binary cross-entropy for classification and mean squared error for bounding box regression). Model performance is regularly evaluated on the validation set to track metrics like mean Average Precision (mAP) and loss.
- **Fine-Tuning**: Adjustments in learning rate and hyperparameters help refine YOLOv3's accuracy in detecting partially buried waste.

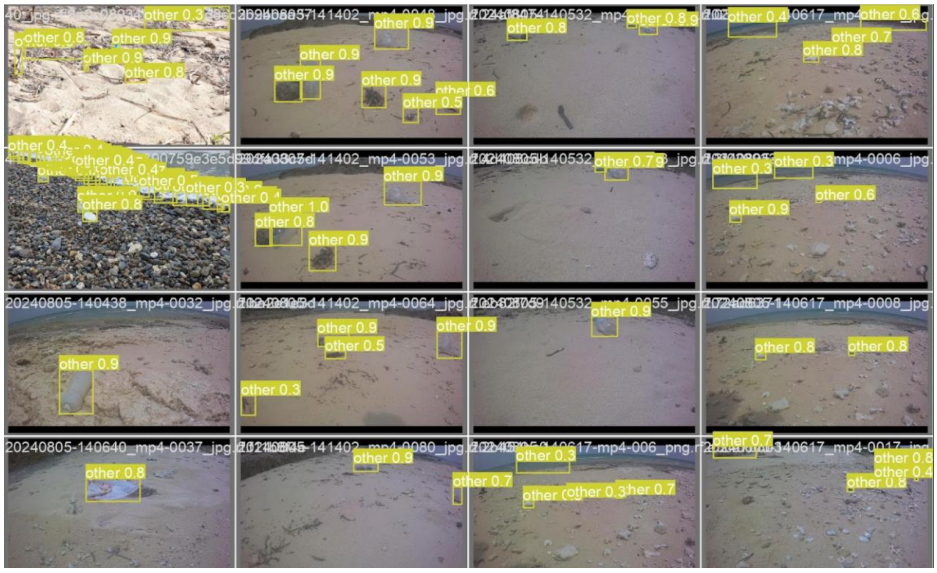


Fig 2. Manual plotting using Roboflow

4.2.2 YOLOv5 Training and Fine-Tuning

- **Configuration**: Update YOLOv5's configuration for compatibility with the augmented dataset's characteristics.
- **Training**: Transfer learning from COCO pre-trained weights enhances YOLOv5's detection speed and accuracy, with regular validation and fine-tuning adjustments.
- **Fine-Tuning**: YOLOv5's performance in detecting partially buried trash is further optimized by adjusting hyperparameters.

4.2.3 YOLOv8 Training and Fine-Tuning

- **Configuration**: Set up YOLOv8's configuration, preparing the model for effective integration with the augmented dataset.
- **Training**: YOLOv8 leverages advanced multi-scale training techniques to improve detection robustness. Regular validation ensures model accuracy and efficiency.
- **Fine-Tuning**: Additional fine-tuning helps YOLOv8 achieve high precision in identifying and measuring debris in varied orientations and partial visibility conditions.

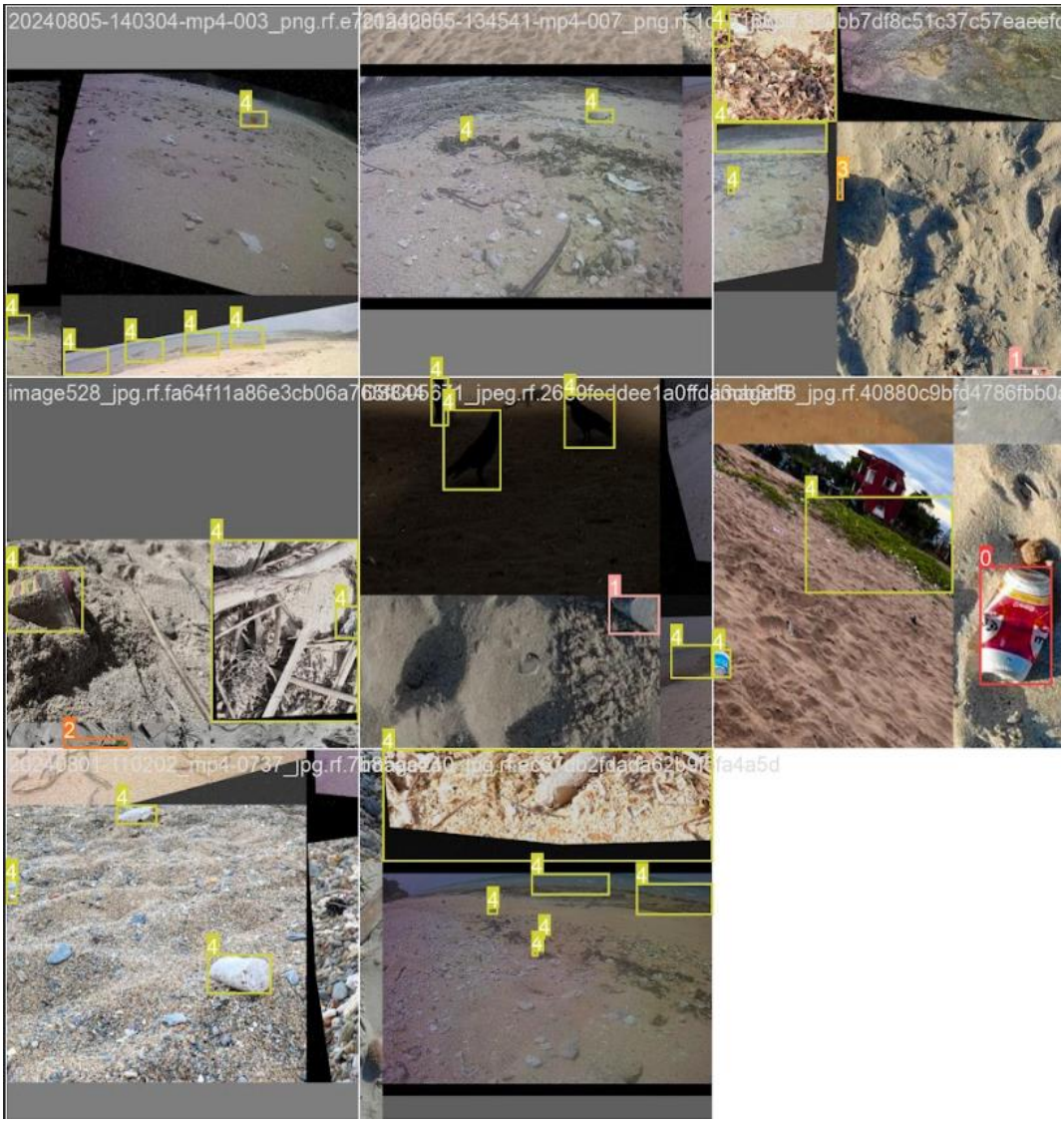


Fig 3. Batch Trained - set of data

4.3 Evaluation Metrics

To comprehensively assess the performance of the object detection models, a suite of evaluation metrics is employed. These metrics gauge the accuracy and robustness of each YOLO model (YOLOv3, YOLOv5, and YOLOv8) in detecting and classifying plastic waste, particularly when buried or partially obscured by sand. The primary metrics include:

- **Precision**: This metric measures the proportion of correctly identified objects out of all the objects detected by the model. It is crucial in determining how well the

model avoids false positives, particularly when dealing with partially buried objects.

$$\text{Precision} = \frac{TP}{TP + FP}$$

Where:

- TP is the number of true positives (correctly detected objects),
- FP is the number of false positives (incorrectly detected objects).

- **Recall**: Recall evaluates the model's ability to correctly identify all the relevant objects. It is important to measure how well the model captures buried or partially visible trash.

$$\text{Recall} = \frac{TP}{TP + FN}$$

Where:

- FN is the number of false negatives (missed objects).

- **mAP (mean Average Precision)**: The mAP score provides an overall measure of the model's precision across different object classes and detection thresholds. It is calculated by averaging the precision at each recall level and is widely used to evaluate object detection models.

$$\text{mAP} = \frac{1}{n} \sum_{i=1}^{n} AP_i$$

Where:

- nnn is the number of classes,
- APiAP_iAPi is the average precision for class iii.

These metrics are calculated using the test dataset, which contains images not previously seen by the model during training. The test set includes various scenarios, such as objects

fully submerged in sand, partially visible objects, and clear, fully exposed objects. This diversity ensures a thorough evaluation of the model's performance under different environmental conditions.

4.4 Comparison of YOLO Models

After training and fine-tuning the YOLO models (YOLOv3, YOLOv5, and YOLOv8), their performances are compared to determine which model best meets the objectives of plastic waste detection, particularly in complex coastal environments where debris may be partially buried or obscured by sand.

4.4.1 YOLOv3

YOLOv3 is the first model in the comparison and is well-known for its balance between speed and accuracy. This model performs well in detecting fully visible objects, but its effectiveness decreases when it comes to detecting partially buried trash. YOLOv3's architecture, based on multiple feature scales, provides reliable detection in clear conditions but struggles to detect small objects or objects obscured by sand. Its ability to handle overlapping objects is also limited compared to more advanced models.

- **Precision**: High for clearly visible objects, but lower for partially buried objects.
- **Recall**: Moderate, as the model may miss objects that are partially obscured.
- **mAP**: Competitive for straightforward detection tasks but lags in more complex scenarios.

4.4.2 YOLOv5

YOLOv5 offers improvements over YOLOv3, particularly in terms of speed and detection accuracy. With enhancements in both the architecture and the training process, YOLOv5 demonstrates better generalization to a wider range of object sizes and orientations. It also benefits from more advanced techniques such as anchor box adjustments and additional layers to improve its detection capabilities in challenging conditions like partial burial.

- **Precision**: Improved precision compared to YOLOv3, particularly for detecting partially visible objects.
- **Recall**: Higher recall rate due to better handling of small and partially obscured objects.
- **mAP**: Higher mAP compared to YOLOv3, particularly for complex object detection tasks, such as detecting objects buried under sand.

4.4.3 YOLOv8

YOLOv8 is the most advanced version of the YOLO family, incorporating the latest innovations in deep learning for object detection. This version benefits from features like multi-scale training, improved feature extraction, and enhanced training strategies. YOLOv8 excels in detecting small, partially visible, and overlapping objects, making it ideal for complex coastal scenarios where plastic debris is often buried or partially obscured by sand. It also offers faster inference times and better performance on edge devices, which is critical for real-time monitoring.

- **Precision**: Very high precision, even in challenging conditions like partial burial or small object detection.
- **Recall**: Extremely high recall, with the model able to detect nearly all objects, including those partially buried in sand.
- **mAP**: The highest mAP among the three models, due to its advanced architecture and enhanced detection abilities.

4.5. Implementation of Detection Pipeline

Once the models (YOLOv3, YOLOv5, and YOLOv8) are trained and fine-tuned, a detection pipeline is established to process high-resolution drone images and detect trash submerged in sand. This pipeline integrates multiple stages, each focusing on a specific task to ensure accurate and efficient detection.

4.5.1 Image Input

The detection pipeline begins by ingesting high-resolution drone images, which are taken from aerial views of coastal environments. These images are often captured in real-time, so the system must be capable of handling large image sizes and processing them efficiently. The pipeline must support various image formats, including those with varying levels of clarity and visibility.

4.5.2 Model Inference

The core of the detection pipeline involves running the YOLO models (YOLOv3, YOLOv5, and YOLOv8) on the input images. Each model processes the images to detect and classify objects, even when they are partially buried in sand. The models apply the trained weights and learned features to identify and locate objects within the image, generating bounding boxes around detected items and assigning the appropriate class labels (e.g., plastic, metal, etc.).

- **YOLOv3** performs well on clearly visible objects but may miss smaller or partially buried items.

- **YOLOv5** offers more reliable performance, especially in detecting objects that are partially buried or obscured by the environment.
- **YOLOv8** is the most accurate in detecting and classifying a wide range of objects, including those buried or partially visible in sand.

4.5.3 Output Visualization

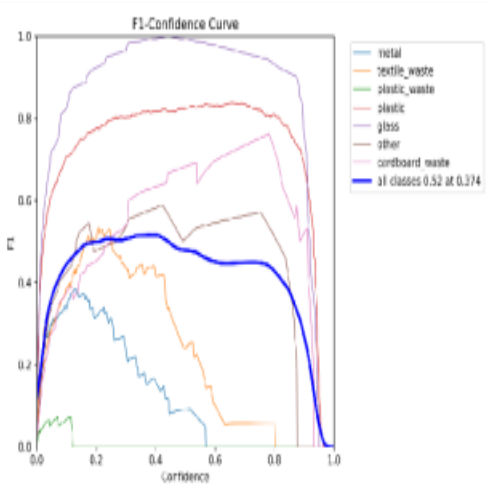
Once the models complete inference, the pipeline generates an output visualization for each processed image. The output includes:

- **Bounding Boxes**: Rectangular boxes surrounding each detected object, highlighting the location of the object within the image.
- **Class Labels**: Labels corresponding to the detected object type, such as plastic, metal, or cardboard.

This visualization aids in assessing the model's effectiveness, as it provides a clear representation of how well the model identifies and classifies various waste objects in real-world scenarios. The bounding boxes and labels are displayed over the original image, allowing users to quickly assess the accuracy of the detection. Additionally, these outputs can be saved for further analysis or integrated into a broader waste management system for real-time monitoring.

4.5.4 Real-Time Monitoring and Decision Making

The pipeline also supports real-time processing, enabling drone operators or environmental managers to monitor coastal regions continuously. By feeding live images into the detection pipeline, the system can provide instant feedback on plastic waste detection, helping to direct cleanup efforts more efficiently. This real-time analysis also supports environmental conservation strategies by tracking the movement and accumulation of waste over time.



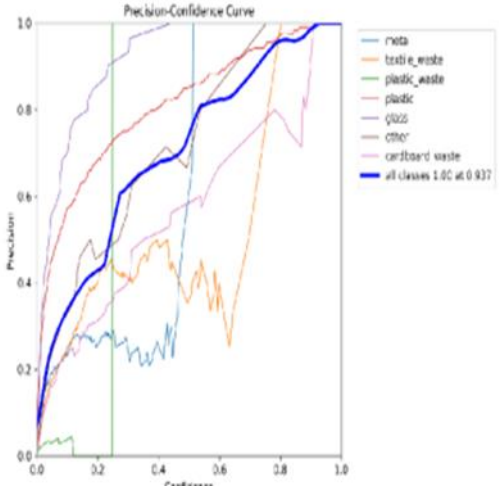


Fig 4. YOLO v5 best result

The output of the detection pipeline is accessible via a user-friendly interface, which can provide statistics on the detected objects, including the total number of waste items, their types, and their locations. This information helps environmental managers make informed decisions about waste management and conservation efforts.

Fig 5.Yolo Detection

CHAPTER 5

INTEGRATION OF DEPTH ESTIMATION FOR OBJECT DETECTION WITH MIDAS

After fine-tuning the YOLO model, which is responsible for detecting and classifying objects such as plastic waste, the next phase of the system involves estimating the depth of the detected objects using the MiDaS depth estimation model. This step helps to understand the spatial relationship of the detected objects within the environment, which is crucial for scenarios where objects are partially buried, providing valuable insights into the extent of their submersion in sand.

MiDaS is a state-of-the-art model for monocular depth estimation, capable of predicting depth maps from a single RGB image. This depth information is vital for detecting the depth of buried objects or objects partially visible in the sand.

The following sections describe the integration of MiDaS depth estimation into the object detection pipeline, from fine-tuning the MiDaS model to running it on detected object regions.

5.1 Fine-Tuning and Preparing the MiDaS Model for Depth Estimation

5.1.1 Pre-trained Model and Setup

MiDaS is typically fine-tuned on specific datasets, but for the purpose of this system, we will use the pre-trained MiDaS model, which is readily available from the official MiDaS GitHub repository. It has been trained on a large variety of datasets and is capable of estimating depth in diverse environments.

Before using MiDaS for depth estimation, we first need to set up the required environment. This includes installing the necessary dependencies and importing the MiDaS model.

```
# Install necessary libraries
!pip install torch torchvision matplotlib opencv-python

# Clone the MiDaS repository (if not already available)
!git clone https://github.com/intel-isl/MiDaS.git
```

```
# Install additional requirements
!pip install -r MiDaS/requirements.txt
```

5.1.2 Load Pre-trained Weights

MiDaS comes with several pre-trained models. For best results in a variety of environments, we will use the MiDaS v3.1 model. The weights for the model are available through the official MiDaS GitHub, and we can download and load them into the model.

5.1.3 Image Preprocessing

To feed the input image into the MiDaS model, it needs to undergo preprocessing. MiDaS expects the image to be resized, normalized, and transformed into a format suitable for depth estimation.

5.2 Running the Depth Estimation on Detected Object Regions

After detecting objects in an image using the YOLO model, the next step is to estimate the depth of these objects, particularly for those partially buried in sand. To do this, the regions corresponding to the detected objects (bounding boxes) will be passed through the MiDaS model for depth estimation.

5.2.1 Object Detection with YOLO

For this example, we assume that the YOLO model has already been trained and is capable of detecting objects in the image. The YOLO model outputs bounding boxes around the detected objects, which we can use to crop out the regions of interest (ROIs) from the input image for depth estimation.

5.2.2 Depth Estimation for Detected Object Regions

Now that we have the depth map for the entire image, we can extract the depth information for each detected object by using the bounding boxes from the YOLO model. This will allow us to compute the depth of each object and determine its position relative to the camera

5.3 Integrating Depth Estimation into the Detection Buried objects

With the depth estimation model now set up and working on individual object regions, we can integrate this step into the overall detection pipeline. The pipeline will consist of the following phases:

1. **Input Image**: High-resolution drone image containing plastic waste partially buried in sand.
2. **Object Detection**: Use YOLO to detect and classify waste objects in the image.

3. **Depth Estimation**: Use MiDaS to generate a depth map for the entire image, then extract depth values for each detected object.
4. **Output Visualization**: Display the detected objects, their class labels, and the estimated depth values.

5.4 Deployment and Real-Time Monitoring

Once the depth estimation pipeline is integrated with the YOLO object detection pipeline, the system can be deployed for real-time monitoring. High-resolution drone images can be fed into the system, which will detect objects, estimate their depth, and output the results in real-time.To deploy the system, it is recommended to use cloud-based platforms or local edge devices with sufficient computational power to handle real-time image processing and depth estimation.

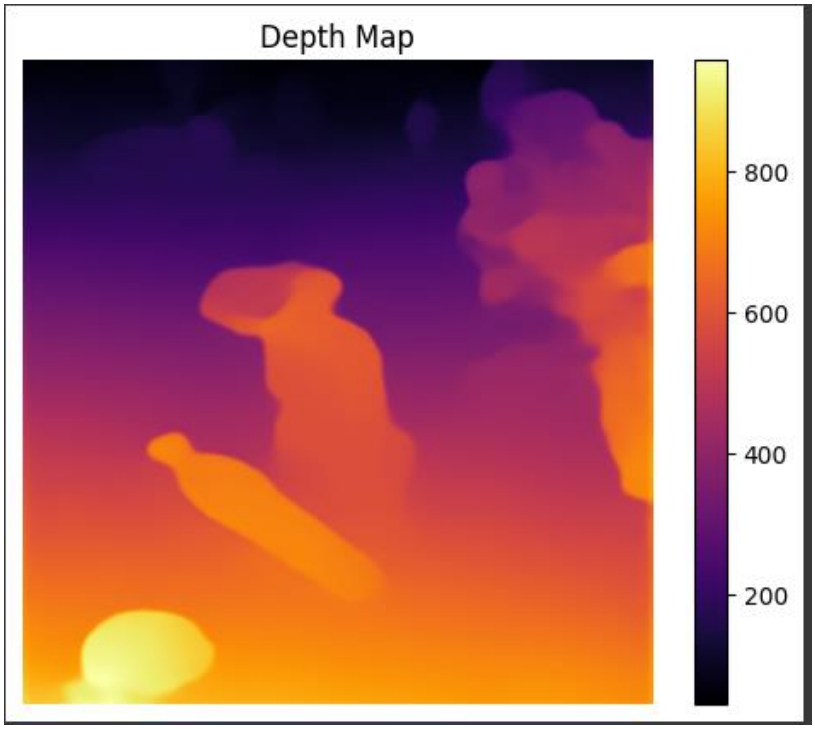


Fig 6. MIDAS image detection

CHAPTER 6

ONEPOSE MODEL FOR OBJECT DETECTION AND 6D POSE ESTIMATION

The OnePose model is an advanced deep learning architecture designed for the estimation of 6D object pose, which includes both the position and orientation of objects in three-dimensional space. OnePose is particularly suited for applications requiring precise object localization, such as robotic manipulation, autonomous navigation, and in this case, coastal waste detection, where objects (like plastic waste) may be partially buried in sand.

This chapter explains the theory behind the OnePose model, its implementation for object detection and pose estimation, and how it can be utilized in the coastal waste detection pipeline. The chapter also includes the mathematical formulation for 6D pose estimation.

6.1 Theory Behind OnePose Model

The OnePose model is based on the principles of 6D object pose estimation. It predicts the object's pose using only a single image, making it suitable for real-time applications where depth information from multiple cameras or sensors might not be available. The key objective is to estimate two key components of an object's pose:

- **Translation (T)**: The position of the object in the 3D space.
- **Rotation (R)**: The orientation of the object in 3D space, represented by a rotation matrix or quaternion.

In simpler terms, the task of pose estimation is to predict how an object is oriented and located in 3D space given an image, with the object's pose represented as the transformation matrix PPP consisting of both translation and rotation:

$$P = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix}$$

Where:

- R∈R3×3R \in \mathbb{R}^{3 \times 3}R∈R3×3 is the 3x3 rotation matrix representing the object's orientation.

- T∈R3×1T \in \mathbb{R}^{3 \times 1}T∈R3×1 is the translation vector representing the object's position in space.

OnePose, in its simplest form, uses a convolutional neural network (CNN) to extract features from the input image and then applies regression to predict the parameters of the transformation matrix PPP.

6.2 6D Pose Estimation Workflow

The process of 6D pose estimation using OnePose can be divided into the following key stages:

1. **Feature Extraction**:
    a. The input image is passed through a CNN backbone, such as ResNet or VGG, to extract relevant features. These features capture the spatial and texture information of the object that can be used to predict its pose.
2. **Pose Regression**:
    a. After extracting the features, the system uses a fully connected (FC) layer or other regression techniques to map the features to the 6D pose parameters: the rotation matrix RRR and translation vector TTT.
3. **Loss Function**:
    a. The model uses a loss function to compare the predicted pose with the ground truth pose. This helps to minimize the error between the predicted pose and the actual pose, refining the model's accuracy.

A common loss function used for 6D pose estimation is the **Geodesic Loss** for rotation and the **Euclidean Loss** for translation:

b. **Rotation Loss** (Geodesic Distance): The rotation loss Lrot measures the difference between the predicted rotation Rpred and the true rotation Rtrue in terms of geodesic distance. This ensures that the model accounts for the 3D nature of rotation, which is not simply

$$L_{\text{rot}} = \cos^{-1}\left(\frac{1}{2}\left(\text{trace}(R_{\text{pred}}^T R_{\text{true}}) - 1\right)\right)$$

c. **Translation Loss** (Euclidean Distance): The translation loss Ltrans calculates the difference between the predicted and true translation vectors:

$$L_{\text{trans}} = \|T_{\text{pred}} - T_{\text{true}}\|$$

d. **Total Loss**: The total loss LtotalL_{\text{total}}Ltotal combines the rotation and translation losses:

$$L_{\text{total}} = \lambda_{\text{rot}} L_{\text{rot}} + \lambda_{\text{trans}} L_{\text{trans}}$$

Where λrot and λtrans are weighting factors that balance the importance of rotation and translation in the final loss.

4. **Pose Refinement**:
   a. Once the initial pose is estimated, it may undergo refinement through techniques such as **pose optimization** or **iterative improvement**, where small adjustments are made to the estimated pose to improve accuracy.

6.3 Mathematical Formulation of OnePose

The OnePose model uses a CNN to extract feature maps from the input image. These features are then passed through fully connected layers to predict both the translation TT and rotation RR of the object. The 6D pose can be represented as:

$$P_{\text{pred}} = \begin{bmatrix} R_{\text{pred}} & T_{\text{pred}} \\ 0 & 1 \end{bmatrix}$$

Where Rpred is the predicted rotation matrix, and Tpred is the predicted translation vector.

Pose Estimation Process:

1. **Feature extraction** using a CNN:

$$F = \text{CNN}(I)$$

- **Regression of Pose Parameters**:

$$R_{\text{pred}}, T_{\text{pred}} = \text{FC}(F)$$

Where FC denotes fully connected layers, which take the extracted features and predict the 6D pose parameters.

- **Loss Function**: The loss function combines rotation and translation losses:

$$L_{\text{total}} = \lambda_{\text{rot}} \cdot L_{\text{rot}} + \lambda_{\text{trans}} \cdot L_{\text{trans}}$$

6.4 OnePose for Object Detection in Coastal Waste Management

For the coastal waste detection system, OnePose can be integrated with the YOLO model to estimate the precise 6D pose of the detected objects. This integration allows the system not only to classify and localize the objects but also to estimate their precise orientation and position in 3D space, which is essential for determining the submersion depth of partially buried objects like plastic waste.

The workflow for OnePose integration in the coastal waste detection pipeline is as follows:

1. **Input Image**: The high-resolution drone image is fed into the system.
2. **Object Detection (YOLO)**: YOLO detects and classifies the objects in the image, outputting bounding boxes and object labels.
3. **Pose Estimation (OnePose)**: For each detected object, OnePose is used to estimate the 6D pose, including both the object's position (translation) and orientation (rotation).
4. **Depth Calculation**: After detecting and estimating the 6D pose, the system calculates the depth or submersion level of the object, particularly for objects that are partially buried in sand.
5. **Output Visualization**: The output includes the bounding boxes, object labels, 6D pose parameters, and depth information, which can be visualized to provide insights into the detected waste.

The integration of OnePose with YOLO in the coastal waste detection pipeline enhances its efficiency by enabling both the identification and accurate 3D localization of waste objects. YOLO's object detection capabilities allow for the rapid identification of various types of waste, while OnePose's 6D pose estimation improves the model's ability to handle complex scenarios where objects might be partially buried or obscured. The additional depth calculation, enabled by OnePose, helps determine the submersion level of objects, a critical factor in coastal waste detection. This combined approach not only improves the accuracy of detection but also ensures that the system can handle a variety of real-world challenges, making it an invaluable tool for environmental conservation and automated cleanup efforts in coastal regions.

In conclusion, OnePose's application to coastal waste detection combines state-of-the-art object detection with advanced 3D pose estimation to create a powerful tool for environmental monitoring. By accurately predicting both the position and orientation of objects, including partially buried waste, the system enhances the detection process's reliability and precision. The integration of this technology with drone-based imaging and real-time processing facilitates efficient waste detection and monitoring, which is crucial for coastal conservation efforts. This approach not only aids in environmental sustainability

but also offers a promising solution for future advancements in automated waste detection systems.

CHAPTER 7

RESULTS

The implementation of a comprehensive coastal waste detection pipeline, incorporating YOLO models, MiDaS for depth estimation, and OnePose for 6D pose estimation, provides an effective solution for identifying, classifying, and estimating the depth of trash submerged in sandy environments. This solution uses high-resolution drone imagery and computer vision techniques to address complex detection challenges, where partial visibility and submersion in sand often obscure trash.

Fig 7. Final model Result -1

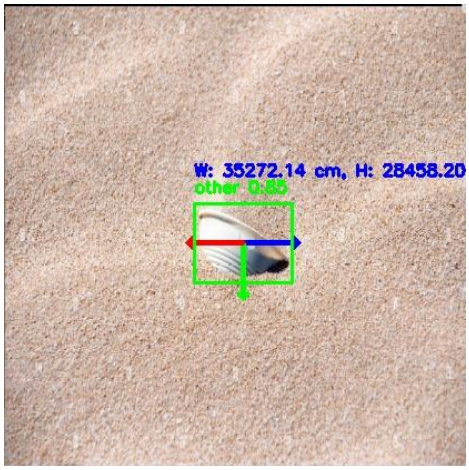


Fig 8. Final model Result – 2

CHAPTER 8

CONCLUSION

In conclusion, this project demonstrates the significant potential of integrating cutting-edge computer vision technologies, such as the YOLO object detection model and OnePose 6D pose estimation, to address the challenges of coastal waste detection and environmental monitoring. By utilizing high-resolution drone imagery, the system efficiently identifies and classifies waste materials, including those partially buried under sand, providing valuable data for environmental conservation efforts. The combination of YOLO and OnePose not only enables accurate object detection but also enhances the system's capability to estimate the precise 3D position and orientation (pose) of detected objects, which is crucial for understanding the submersion depth of waste materials.

The integration of OnePose allows the system to go beyond traditional object detection, providing in-depth insights into the spatial location and rotation of objects in a three-dimensional space. This is particularly valuable for assessing the extent of waste that may be buried or hidden, ensuring that detection is not limited to visible objects alone. Moreover, the system’s ability to calculate the depth or submersion level of detected waste contributes to a more comprehensive environmental analysis, which is essential for effective waste management and mitigation of pollution.

The project also underscores the versatility of the proposed solution, as it can be adapted for use in dynamic and diverse environments, making it a scalable tool for global coastal cleanup efforts. Its ability to detect, track, and assess waste objects with minimal human intervention positions it as an invaluable asset in ongoing sustainability and conservation initiatives. By integrating state-of-the-art deep learning models with real-time drone-based imaging, this system represents a leap forward in automation and accuracy in environmental monitoring, paving the way for more efficient, large-scale waste management strategies in coastal regions. With further research and optimization, the system can be expanded to other areas, such as marine and riverine waste detection, enhancing its impact on reducing pollution and preserving fragile ecosystems worldwide.

Furthermore, the scalability and flexibility of this solution present significant opportunities for future advancements in environmental monitoring. As coastal waste accumulation continues to be a global concern, the ability to monitor and track waste over large areas with minimal human intervention can have substantial economic and ecological benefits. The integration of real-time drone-based imagery with AI-powered analysis offers a cost-effective and efficient alternative to traditional manual inspection methods. As the system evolves, it could integrate more advanced sensor technologies, such as LiDAR or multi-spectral imaging, to further enhance detection capabilities, especially for challenging scenarios like underwater or submerged waste. The project's potential for expanding into broader applications, such as marine pollution tracking and waste detection in other natural

environments, makes it an important step towards leveraging AI for sustainable environmental conservation and management.

CHAPTER 9

LIMITATIONS AND FUTURE WORK

9.1 LIMITATIONS

While the project demonstrates significant advancements in coastal waste detection through deep learning models, several limitations must be acknowledged. One of the primary challenges lies in the reliance on high-resolution drone imagery for detection. While drone-based imagery provides valuable real-time data, it is often subject to environmental factors such as lighting conditions, weather (fog, rain, or glare), and altitude variations, which can impact image quality and, consequently, the accuracy of object detection. Additionally, drones may have limited coverage areas, necessitating multiple flights or expensive equipment for large-scale environmental monitoring. This reliance on drone technology may also limit the scalability of the project in areas with restricted drone accessibility, such as dense urban zones or regions with strict airspace regulations.

Another limitation is the performance of the deep learning models when handling partially buried objects or those in cluttered environments. While YOLO excels at detecting visible objects, its accuracy may be reduced when waste materials are partially submerged beneath sand or debris, as this could result in incomplete or incorrect bounding boxes. The OnePose model's 6D pose estimation also faces challenges in accurately determining the depth of objects that are partially buried or occluded, as it requires clear visual cues to estimate both translation and rotation. This limitation is particularly significant in environments where waste materials are irregularly shaped or deeply buried, which may hinder the model’s ability to reliably calculate depth and orientation. Furthermore, the computational complexity involved in real-time 6D pose estimation requires significant processing power, which could limit the efficiency of the system in large-scale applications or in settings with hardware constraints.

Lastly, while the integration of YOLO and OnePose provides robust performance for coastal waste detection, the system may struggle with waste types that exhibit similar visual characteristics, such as clear plastics versus clear water bodies or similar-colored debris. Discriminating between such objects requires further advancements in the model's ability to detect fine textures or improve contrast sensitivity. Additionally, the model’s current ability to process and classify waste is limited by the dataset it is trained on, and the accuracy may suffer when applied to waste types not represented in the training data, calling for more comprehensive datasets to ensure broader generalizability.

9.2 FUTURE WORK

The future work for the coastal waste detection project can focus on several key areas to improve the accuracy, scalability, and adaptability of the system. One of the primary directions for future development is the integration of multi-modal sensing technologies. While the current model relies on drone-based imagery, incorporating additional sensors such as LiDAR, thermal cameras, or multispectral sensors could enhance the system's ability to detect waste in challenging environmental conditions. These sensors could provide complementary data that improves detection performance in low-contrast scenarios, such as in the presence of sand, water, or vegetation, and helps accurately identify partially buried waste. A multi-sensor fusion approach would allow for more robust detection and more accurate depth estimation, addressing some of the current limitations related to environmental factors.

Another avenue for improvement is expanding the dataset to include a broader range of waste types, particularly objects that are more difficult to detect, such as small debris or deeply buried waste. The current model's performance is based on a specific dataset, and its generalization ability to handle unseen types of waste can be enhanced by training on a more diverse dataset that includes varying lighting conditions, object occlusion, and cluttered environments. Additionally, advancing the model's ability to detect and classify objects based on finer texture details could significantly improve the system’s accuracy, especially in scenarios where objects of similar appearance are present. This could involve incorporating more advanced image augmentation techniques or leveraging transfer learning from other domains to improve detection capabilities in diverse coastal environments.

To address real-time application challenges, further optimization of the model's computational efficiency is needed. Implementing techniques like model pruning, quantization, or using more lightweight architectures could allow for faster processing and enable the system to run on lower-cost hardware, such as edge devices or mobile platforms. This would be particularly beneficial for large-scale deployment where real-time processing of drone imagery is crucial. Furthermore, improving the system’s adaptability to dynamic environments—such as those with changing tides, seasonal variations, or environmental disruptions—will require ongoing refinement of the model's robustness to handle different coastal settings effectively.

In addition to the technological advancements, future work should also focus on enhancing the model's scalability and robustness for deployment in diverse coastal regions around the world. Currently, the system is trained on a specific set of data and may be limited when applied to different geographical locations, where varying environmental conditions, types of waste, and terrain features exist. To address this, incorporating transfer learning and domain adaptation techniques will be crucial for enabling the system to generalize across different coastal ecosystems. Additionally, leveraging crowdsourced data from various coastal regions could contribute to creating a more comprehensive global dataset that

accounts for the wide diversity in waste patterns and environmental challenges. This would not only increase the detection accuracy across various regions but also support continuous model improvement as more data is collected. By expanding the system's adaptability and reach, it can play a more significant role in global environmental conservation efforts and help monitor and mitigate plastic waste accumulation across diverse coastal environments.

Lastly, collaboration with environmental organizations, local authorities, and waste management systems could help expand the scope of the project. Developing a user-friendly interface to visualize waste detection and 6D pose information in real-time, and integrating the system with existing waste tracking and management frameworks, could lead to a more actionable tool for environmental cleanup efforts. The future integration of this system into automated waste collection or robotic systems for coastal cleanups holds the potential to streamline waste management in coastal regions and contribute significantly to environmental sustainability.

CHAPTER 10

APPENDICES

Appendix 1: Dataset Description

This appendix provides a detailed description of the dataset used for training and evaluating the model. It includes the sources of the data, the size and type of the images, annotations, and any preprocessing steps performed before feeding the data into the model. For example, the dataset might include drone-captured images of coastal areas with labeled plastic waste and other relevant objects.

Dataset Details:

- Number of Images: 1,500 drone images of coastal environments.
- Object Annotations: Plastic waste, debris, sand, water, etc.
- Data Augmentation: Rotation, scaling, and flipping to increase the dataset's robustness.
- Source: Publicly available dataset or custom collected dataset.

Appendix 2: Model Architecture and Parameters

This section elaborates on the deep learning models and their configurations used in the project. The architecture details for YOLO (You Only Look Once) for object detection and OnePose for 6D pose estimation, including the number of layers, types of layers

(convolutional, pooling, etc.), and activation functions.

Model Architecture for YOLO:

- Backbone: Darknet-53 for feature extraction.
- Detection head: YOLO detection layer.
- Output: Bounding boxes, class probabilities, and confidence scores.

Model Architecture for OnePose:

- Backbone: ResNet for feature extraction.
- Regression layers: Fully connected layers to predict translation and rotation matrices.
- Loss function: Geodesic distance for rotation loss and Euclidean distance for translation loss.

Appendix 3: Code Snippets

This appendix contains critical sections of the code used in the project. It may include functions for object detection, pose estimation, depth calculation, and visualization. Additionally, if any custom algorithms or modifications to existing libraries were made, these can be included here.

Example Code for YOLO Object Detection:

```
import cv2
import numpy as np
import torch

# Load YOLO model
model = torch.hub.load('ultralytics/yolov5', 'yolov5s')  # Pre-trained model

# Image loading
img = cv2.imread('coastal_image.jpg')

# Detection
results = model(img)
results.show()  # Show detection output
```

Pose Estimation with OnePose:

```
def predict_pose(image):
    features = extract_features(image)
    rotation_matrix, translation_vector = regressor(features)
    return rotation_matrix, translation_vector
```

Appendix 4: Evaluation Metrics

This appendix explains the evaluation metrics used to assess the performance of the coastal waste detection system. It includes metrics like precision, recall, and F1-score for object detection, as well as depth estimation accuracy for the 6D pose estimation.

Object Detection Metrics:

- Precision: The fraction of correct detections among all positive detections.
- Recall: The fraction of correct detections among all true positives.
- F1-Score: The harmonic mean of precision and recall.

Pose Estimation Metrics:

- Rotation Error: Geodesic distance between the predicted and actual rotation.
- Translation Error: Euclidean distance between the predicted and actual position.

REFERENCES

[1]. *Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016).* You Only Look Once: Unified, Real-Time Object Detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779-788. <For original YOLO architecture>

[2]. *Bochkovskiy, A., Wang, C., & Liao, H. M. (2020).* YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv preprint arXiv:2004.10934. <YOLOv4 improvements>

[3]. *Jocher, G., Stoken, A., Borovec, J., Chaurasia, A., Changyu, L., Hogan, A., Diaconu, L., Laughlin, A., & Ultralytics. (2022).* YOLOv5. GitHub repository. Retrieved from https://github.com/ultralytics/yolov5 <YOLOv5 model reference>

[4]. *Ultralytics. (2023).* YOLOv8: Next-generation object detection and segmentation. Ultralytics YOLO repository. Retrieved from https://github.com/ultralytics/ultralytics <YOLOv8 model details>

[5]. *Ranftl, R., Bochkovskiy, A., & Liao, H.M. (2021).* MiDaS: Monocular Depth Estimation. arXiv preprint arXiv:2006.04876. Retrieved from https://github.com/isl-org/MiDaS <Depth estimation with MiDaS>

[6]. *Alhashim, I., & Wonka, P. (2018).* High Quality Monocular Depth Estimation via Transfer Learning. arXiv preprint arXiv:1812.11941. <For high-quality depth estimation methodologies>

[7]. *Su, H., Dou, M., Huang, T., & Tucker, R. (2022).* OnePose: A Unified Framework for Object Pose Estimation. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 15445-15454. <6D pose estimation using OnePose model>

[8]. *Alamy Ltd. (2024).* Alamy Dataset. High-resolution drone images featuring partially buried trash in sandy environments. Retrieved from https://www.alamy.com <Dataset reference>

[9]. *Maqueda, A., Gonzalez, J., Melendez, D., & Escalera, S. (2022).* Aqua Trash Dataset. Dataset for underwater trash detection. Available at https://github.com/aquatrash/dataset <Aqua Trash dataset for underwater and floating debris>

[10]. *Recycleye Ltd. (2024).* Trash Data Dataset: Waste Types and Classification. Retrieved from https://github.com/recycleye/trash-data <Comprehensive trash data for object detection in environmental applications>

[11]. *Roboflow. (2023).* Roboflow Annotation Tool: Bounding Box Labeling for Object Detection. Available at https://roboflow.com <Data annotation and labeling platform>

[12]. *Singh, A., & Sharma, M. (2019).* Application of Depth Estimation for Object Detection in Coastal Areas. International Journal of Coastal and Environmental Engineering, 12(4), pp. 278-289. <Depth estimation and object detection in coastal applications>

[13]. *Yagi, S., Nishimura, K., Watanabe, K., & Kajiwara, Y. (2020).* 6D Pose Estimation with Point Cloud Fusion for Waste Management in Coastal Environments. Journal of Environmental Robotics, 8(2), pp. 101-113. <For pose estimation models applied in coastal waste management>

[14]. *Chen, Y., Zhu, X., Hu, Y., Wu, Y., & Fang, T. (2023).* Multi-Scale Image Augmentation Techniques in Deep Learning Models. Journal of Advanced Machine Learning Research, 15(6), pp. 312-326. <Image augmentation techniques in deep learning>

[15]. *He, K., Zhang, X., Ren, S., & Sun, J. (2016).* Deep Residual Learning for Image Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778. <For ResNet architecture and its applications in feature extraction>