Automated Detection of Structural Anomalies Using Object Tracking Techniques

Praveen Venkumahanti
dept. of Artificial Intelligence and Data Science (AI&DS)
GMRIT
Rajam, India
vpraveen0049@gmail.com

Vamsi Cheekati
dept. of Artificial Intelligence and Data Science (AI&DS)
GMRIT
Rajam, India
saivamsicheekati@gmail.com

Manohar Tammisetti
dept. of Artificial Intelligence and Data Science (AI&DS)
GMRIT
Rajam, India
rockervenkat7@gmail.com

Abstract

Civil engineering structures - buildings, pavements, and bridges, among others - are the cornerstones of modern infrastructure. These are exposed to degradation from various environmental conditions, material fatigue, and usage patterns. This often evolves into developing defects over time. Therefore, early detection of these anomalies is critical in ensuring safety, longevity, and economic viability. This is a very important task of civil engineering as detection of defects in time may prevent probable failures. Common anomalies include cracks in buildings, potholes in pavements, corrosion in bridges, and other structural defects. Inspections by the conventional approach are manual, time-consuming, and liable to human mistakes. An automated and efficient approach must be identified to detect these anomalies at an early stage. This project develops an automated system that detects anomalies in buildings, pavements, and bridges in civil engineering structures early, using advanced image segmentation techniques. We first collect images from the web and then create a custom dataset to train and evaluate segmentation models like YOLO and Detectron2. Such models will be fine-tuned for the more accurate identification and segmentation of structural anomalies so that maintenance interventions can be implemented on time and with accuracy. The implementation of image segmentation will not only increase the precision and efficiency of defect detection but will also improve the safety and durability of civil engineering structures. This project aims to provide infrastructure management with an extentable solution, given the automation of the inspection process in order to detect potential issues at an early stage and also extend the lifetime of such structures.

Keywords—*Structural Anomaly Detection, Object Tracking Techniques, Automated Inspection, YOLO, Detectron2, Structural Defect Identification.*

I. INTRODUCTION

In the fast-growing modern urban environment, this rapid growth of population and urbanization increases a vast amount of waste generation, creating severe challenges towards hygiene, environmental sustainability, and public health. Effective waste management has become the backbone of sustainable urban development. Several such challenges include monitoring public dustbins in real time to ensure on-time collection and avoid overflowing that makes the place utterly unsanitary and contributes to environmental hazards.

In traditional methods of waste collection, the predetermined schedules are prone to much inefficiency in terms of unnecessary collection trips or overlooking overflowing dustbins. Such inefficiencies waste resources and reflect poorly on the general environment and public perception of cleanliness within the urban center. In this context, the integration of AI into a waste management system is promising towards addressing these inefficiencies.

II. This paper puts forth the comprehensive system for automated dustbin monitoring using the most advanced techniques in object detection, tracking, classification, and segmentation. The system proposed above is designed to provide insights about the present usage of dustbins into the municipalities and waste management organizations so that they can optimize the collection schedules, reduce costs, and promote sustainable urban environments. Efficient waste management is crucial to maintain cleanliness, hygiene, and livability in the urban area. Further, the high pace of rapid urbanization and population growth are generating an increase in waste. Traditional collection systems are proving inappropriate in these situations. The demand for smarter, data-driven solutions ensures timely and effective management of waste in such growing scenarios.

III. Fixed schedules for collecting waste, therefore, often become decoupled from actual utilization levels of dustbins. It leads to situations where collection vehicles visit underutilized dustbins consuming fuel and man-hour, but the other areas that require attention are left unattended with dustbins full and overflowing. Such imbalances inflate not only the operational costs but also the carbon footprint associated with waste management activities.

IV. Overflowing dustbins pose a serious health and environmental hazard. They attract pest infestation, spread bad smells, and might cause the diffusion of diseases. Improperly managed waste contributes to pollution, can block drainage systems, and even hurt biodiversity. Such scenes degrade the quality of life in urban places and tarnish the image of the city. Hence, traditional ways mainly rely on manual inspections and periodic checks, failing to capture real-time data regarding the status of dustbins. This renders the waste management systems incapable of responding dynamically, leading to delayed actions and inefficient resource allocation.

V. LITERATURE SURVEY

[1] The paper tracks the advance in object detection starting with the earliest methods-those were the Viola-Jones framework for face detection and HOG, based on a pixel histogram descriptor. It focuses particularly on two-stage detectors, such as RCNN and Fast-RCNN, which rely on complex architectures and use selective region proposals, and one-stage detectors, like YOLO, which focus on faster inference and process all spatial regions at once. The review deals with the architectural successors of YOLO, which discuss the improvement in the design and optimization techniques leading to a better accuracy of detection. and efficiency. The paper covers popular datasets such as Pascal VOC and discusses evaluation metrics like mean Average Precision (mAP) and Average Precision (AP), which are crucial for assessing the performance of object detection algorithms.

[2] The paper discusses various methods on Multiple Object Tracking and points out the distinction between tracking-by-detection frameworks and online methods. Some methods use all object candidates globally whereas others are particularly focused on real-time processing using current information of frames. Global methods are usually optimization techniques, such as network flow and Bayesian filtering. Other online methods rely on two-way cost calculations across appearance and motion that are particularly best used in real-time applications. It is hence essential to compute the match costs based on the appearance and motion cues. Traditional approaches are RGB colour histograms and learned features. Deep learning techniques are used in new approaches for increased accuracy. The paper refers to the advancement in deep learning, like utilizing recurrent neural networks to maintain similarity scores and discriminative learning methods for appearance matching. These innovations aim at improving tracking performance while working on the challenges of false positives and identity switches.

[3] The paper states that the superiority of deep neural networks over other algorithms is the reason for their preference in many classification applications. However, the training of these networks is a time-consuming process and can take days or weeks on large datasets. It talks of transfer learning-the idea of reducing training time by using knowledge from pre-trained models-and its performance improves even in the context of small datasets, thus is worth using in image classification applications. The paper refers to previous work where transfer learning had been used for classifying maturity in papaya and video scenes in sports. It also mentions pre-trained networks like VGG16 and Wide Residual Networks that have been applied for land use classification. Three pre-trained models-MobileNet V2, ResNet50, and VGG19 are studied through metrics such as accuracy, precision, recall, and F1-score, which helps in understanding the performance of these models on a custom dataset.

[4] This paper conducts research on the integration of advanced technologies to improve the detection and localization of radiological and nuclear materials in urban settings. Indeed, this study focuses much attention on networked detector systems, which are needed for identifying radiological threats in challenging urban environments. These systems struggle to interpret alarms based on the need for situational awareness as specified by spectroscopic data only. The contribution of the paper lies in innovative use of LiDAR technology in conjunction with streaming video for real-time object detection and tracking. In this concept, such fusion enhances the improvement of situational awareness and detection sensitivity compared to traditional methods. A new important contribution of this work is the formulation of a real-time analysis pipeline correlative to object

trajectories in spectroscopic gamma-ray data. This capability allows discrimination between source-carrying and non-source-carrying objects, thereby making radiological alarms more effective. The authors provide a quantitative performance assessment of their object detection and tracking methodologies. This assessment characterizes the source-object attribution capabilities of both video and LiDAR systems, demonstrating their effectiveness in a mock urban environment. It discusses a contextual-radiological data fusion methodology that utilizes tracking information to improve detection sensitivity.

[5] This paper titled "Unified Model Based on Reinforced Feature Reconstruction for Metro Track Anomaly Detection" surveyed literature on the evolution and the current state of unsupervised anomaly detection techniques in the safety context of metro tracks. Its emphasis lay in the fact that with the scarcity and uncertainty of abnormal samples, there is increasing interest in unsupervised methods of anomaly detection. Such methods are very important when one has limited labelled data in industrial applications. The existing methods can broadly be classified into two categories: Reconstruction-based Methods: It reconstructs normal data patterns and anomaly identification based on reconstruction errors. Feature Embedding-based Methods: It embeds features in a lower-dimensional space to differentiate between normal and abnormal samples. It identifies the significant challenges that current unsupervised methods face; traditional methods often depend on one model per category or scene, which is insufficient for the complex and diversified environments of metro tracks leading to poor stability and performance. Feature-based methods also face challenges when handling multistage features from pretrained models that contain redundant information and noise, causing interference in the effective reconstruction of features and anomaly detection. The reinforced feature reconstruction-based anomaly detection network to overcome these challenges. The proposed model includes the efficient channel feature reinforcement module and a special loss function to improve feature representation and reduce noise interference.

[6] The paper "Rethinking Motion Estimation: An Outlier Removal Strategy in SORT for Multi-Object Tracking with Camera Moving" is presented with a broad overview and scope of already published literature on the multi-object tracking (MOT) domain and motion estimation. The improvement trends of object detection technologies were very rapid, indicating a progressive development of this area. This has significantly enhanced the performance of multi-object trackers; the developments henceforth have reduced complex tracking algorithms instead leading to more effective tracking solutions. Primarily, their work focuses on motion-based multi-object tracking. Integrating object detection into trajectory prediction is essential in effectively tracking the movement of multiple objects in dynamic environments. The paper discusses the Simple Online and Real-Time Tracking (SORT), which is one of the common methods used in MOT. SORT utilizes the dual-phase Kalman Filter for estimating object state. The authors note that existing SORT-like algorithms commonly fail to handle camera motion appropriately and therefore tend to perform suboptimally in dynamic scenarios. Therefore, they introduce Outlier Removal-based SORT, or OR-SORT, which modifies the traditional SORT framework with a triple-phase Kalman Filter to separate the object's motion state transition model into parts. This splitting is essential for better precision on tracking in camera movement. It also shares the method of implementing an outlier removal algorithm using Mixed Integer Linear Programming (MILP). This outlier removal algorithm aims to enhance the robustness of the camera motion estimation, which plays a significant role in the performance of multi-object tracking under highly dense scenarios with moving cameras.

[7] As presented in the paper "A Novel Method for Road Anomaly Objects Detection in the Traffic Environment with Multi-Mechanism Fusion," there exist various approaches used in road anomaly detection, advancing from traditional methods to advanced deep learning techniques. Initially, it was through sample division of known datasets that enabled the process of obtaining features and classification using shallow classifiers. These methods were somewhat limited in their detection and classification capabilities; these were not as effective in scenarios. It reports a massive shift in recent years towards deep learning methods. Convolutional Neural Networks (CNNs) have appeared as the most promising tools for feature extraction and classification, which can then help achieve more accurate road anomalies' detection. This transition marks an important improvement from earlier shallow classifier methods. The proposed method in this paper combines an assistant method and a resynthesis method, which work together to enhance the accuracy of detection while reducing noise. Another improved feature is the addition of a postprocessor using linear clustering to segment road images into super pixels. It improves the localization accuracy and reduces the false positives achieved due to the anomaly scoring of each super pixel. The proposed method is also compared with the other existing methods, such as FlowEneDet, DaCUP, to point out their strengths and weaknesses. For example, even though the FlowEneDet shows high accuracy, it generates a large number of false positives; this problem is addressed in the proposed method. The efficiency of the proposed method is verified by conducting experiments on datasets such as the Obstacle Track and Lost and Found datasets.

[8] The paper "Automatic Object Tracking and Segmentation Using Unsupervised SiamMask" gives a detailed overview of the state-of-art in object detection, tracking, and segmentation. It emphasizes that an effective tracking system consists of two main components, namely, object detection and object tracking. It thus focuses the need for progress in both to gain overall performance enhancements. Various methods of object detection have been discussed. The following include YOLO (You Only Look Once): Known for its accuracy in detecting smaller objects but struggles in larger ones. All these variants of YOLO use CNNs of different sizes to return bounding boxes around detected objects. Single Shot Multi Box Detector (SSD): This algorithm also detects objects in a single forward pass and uses multiple independent logistic classifiers for class prediction and is trained using binary cross-entropy loss. Detectron2: Released by FAIR, this framework is designed for object detection and

segmentation, achieving real-time speed with high accuracy. It discusses the problems that occur with object tracking, mainly in real-time scenarios. It refers to the S-Siam framework, pointing out issues with a loss of accuracy due to camera jitter as well as with a small and high-speed-moving object. A list of datasets used for benchmarking object detection and tracking algorithms appears such as YouTube-VOS and VOT2020. YouTube-VOS is known for its larger amounts of high-resolution video and manual annotations, whereas VOT2020 emphasizes on both short- and long-term tracking.

[9] The paper "DMFF-YOLO: YOLOv8 Based on Dynamic Multiscale Feature Fusion for Object Detection on UAV Aerial Photography" addresses the challenges of object detection in aerial images captured by drones. These factors complicate the detection process, leading to suboptimal performance of traditional object detection models, which are primarily trained on natural images. Mainstream object detection models, including the R-CNN series, YOLO series, and DETR series, often struggle with aerial target detection due to their training on natural scenes. This results in poor performance when applied to the specific characteristics of drone imagery. To tackle the multiscale issues in aerial imagery, researchers have developed various feature fusion techniques. For instance, the FPN-PAN structure in YOLOv5 combines deep semantic features with shallow detail features, while attention-based algorithms have been proposed to enhance the network's capability to extract multi-scale data. Several innovative methods have been introduced to improve feature extraction for small and densely distributed aerial targets. The proposed method outperforms existing algorithms, particularly in detecting small and dense targets, while also reducing the number of parameters and computational load. It concludes that while many methods have been proposed to address the challenges of aerial image detection, existing solutions still fall short in effectively handling the unique characteristics of drone imagery.

[10 Some of the main areas that capture the scope of the literature on detecting and tracking defects in drainage pipelines revolve around Object detection and tracking: The two fundamental operations within artificial intelligence visual techniques are object detection and tracking. The former refers to finding abnormalities, such as defects, in one image, while the latter compares consecutive frames in order to maintain consistency of identified defects through time. Such tasks hold an importance of their own to develop reliable inspection models for the defects in pipelines, allowing for the instant detection and accurate monitoring of urban infrastructure. The huge amount of visual data sensed through technologies such as CCTV or drones requires the use of intelligent techniques for automated defect identification. The improved performance aims at boosting the efficiency and accuracy of defect detection processes; underlines the role of deep learning in defect analysis, especially in the process of automated detection and classification of several types of defects. While much has been achieved so far, the quantification of multiple defect categories is still problematic, suggesting a future research direction in this area. It proposes that active methods of inspection are important for the early detection of pipeline defects, thus preventing environmental pollution and damage to infrastructure systems. This is contrary to old-fashioned passive maintenance methods, which incur increased costs and dangers. The proposed system integrates detection and tracking algorithms into an information management platform. This will facilitate the intelligent identification and reporting of defects in pipelines. This integration streamlines the inspection process and improves overall efficiency in defect management.

[11] Literature Survey of Multiple Object Tracking (MOT) The paper "Multiple Object Tracking with Attention to Appearance, Structure, Motion and Size" presents a comprehensive review of existing methodologies in the field of Multiple Object Tracking (MOT). MOT methods can be broadly categorized into tracking-by-detection frameworks, where object candidates are identified using detectors. Global methods often employ techniques like network flow optimization, graph-based clustering, and Bayesian filtering, while online methods focus on real-time applications by using pair-wise costs for matching detections to tracks. It highlights the importance of calculating matching costs based on appearance and motion. Traditional methods include RGB colour histograms and learned features, while motion-based matching costs consider expected positions and spatio-temporal distances. It covers recent advancements in deep learning, where methods have been developed that utilize recurrent neural networks (RNNs) for calculating similarity scores based on appearance, motion, and social interactions. Additionally, it notes that while deep learning methods can achieve high accuracy, they often require significant computational resources and training data. It indicates a need for simpler, more efficient methods that can balance accuracy and real-time performance, especially for applications in robotics.

[12] Survey of Deep Learning Techniques for Real-Time Multi-Object Tracking This paper on "Analysis Based on Recent Deep Learning Approaches Applied in Real-Time Multi-Object Tracking: A Review" presents a thorough literature survey with regard to state-of-the-art, developing techniques in deep learning approaches towards online MOT. The literatures reviewed were stratified into four primary themes. This theme of Online MOT Based Detection discusses methods that enhance the detection quality and improve the association of detected objects with their respective tracks. Real-Time MOT with High-Speed Tracking and Low Computational Costs High-speed tracking with low computational costs is achieved by approaches highlighted in this section: approaches that reach efficient tracking while maintaining low computational demands, crucial for real-time applications. Modelling Target Uncertainty in Online MOT The authors reviewed approaches dealing with uncertainty in tracking, especially in challenging environments where object appearances and movements become very different. Deep Convolutional Neural Network (DCNN), Affinity, and Data Association: This theme focuses on the integration of DCNNs with affinity measures to improve data association processes in tracking. As part of the literature survey, there were substantial

challenges left unresolved in this field of endeavor, such as problems of losing objects in very cluttered scenes, varying motion, and appearance change. These are often the cause of a drop in the accuracy of tracking performance. It is useful for the researchers and practitioners working in computer vision to gain the thrust in the current techniques and challenges that remain in real-time multi-object tracking.

[13Literature Survey of the Paper The paper presented a directional-edge-based object tracking system, exploiting ideas and methodologies offered by current literature. The system was fundamentally motivated from biological principles, particularly the way animals apply directional edges in visual tracking. This further underline the need to comprehend basic biological mechanisms while putting down efficient tracking algorithms. It uses a directional-edge-based image feature representation algorithm that has been proven robust in many image recognition applications. This is very important for representing object features in more challenging conditions, such as changes in illumination or deformation of an object. Some aspects discussed involve putting online learning into tracking in order to enhance the performance tracking. This technique adapts the system to the changes in the target object's appearance; thus, it is a great advancement of traditional tracking methods which do not hold for such variations. It highlights the computational efficiency of its approach, which is particularly important for real-time applications on hardware platforms like FPGA. It also references previous works on dedicated VLSI chips for efficient directional edge detection, indicating a trend towards hardware-friendly algorithms that can be integrated into larger systems. The paper proposes potential solutions to these issues, indicating a comprehensive understanding of the current state of research in the field. Overall, the paper reflects a thorough examination of existing methodologies and highlights the contributions of the proposed system in advancing the field of object tracking.

[14] Literature Survey of Multi-Object Tracking The paper "Online Multi-Object Tracking with Visual and Radar Features" broadly represents the literature in the existing field, multi-object tracking (MOT), with greater emphasis on radar-based and vision-based methods. Radar systems consist of spatial detection information, such as range and bearing. Still, it encounters problems like mixed signals from multiple objects and clutter. Several data association methods have been developed to connect measurements to associated tracks. These include simple greedy methods such as nearest neighbor and strongest neighbor associations, more sophisticated techniques such as Joint Probabilistic Data Association (JPDA) and Multiple Hypothesis Tracking (MHT), which were specifically designed to handle the complexities of joint track-to-measurement assignments. To reduce computational complexity, Linear Multi-target Integrated Probabilistic Data Association (LMIPDA) has been introduced, along with Sequential Monte Carlo (SMC) methods to manage nonlinear dynamics. Vision-based tracking has evolved significantly, with methods categorized into batch and online tracking. Batch methods utilize global associations across entire frames, yielding better results but lacking real-time applicability. Affinity evaluation between tracks and detections has been given utmost importance, and hence developed several object affinity models using appearance, motion, and shape cues. Also, recent advancements in deep learning culminated in the development of deep learning-based affinity models that improve model discriminability but are requiring large training data and also computational resources. It introduces an amplitude affinity model for vision-based MOT, arguing that it can improve tracking accuracy while being low in complexity. This model is compared against recent deep learning methods, showing competitive tracking accuracy on a challenging dataset. It highlights the evolution of MOT techniques, challenges faced in both radar and vision-based systems, and innovative approaches proposed in the paper to further enhance tracking performance.

[15] The literature review conducted in paper "FFTransMOT: Feature-Fused Transformer for Enhanced Multi-Object Tracking" provides a broad overview of existing methodologies in the field of MOT. MOT is considered as an approach in terms of algorithms aimed at tracking a multiple number of objects in video sequences and capturing their positions, velocities, sizes, and movement patterns. Such foundational insight lays a foundation for various tracking methodologies. It focuses on conventional approaches to tracking, of which theHistograms of Oriented Gradients (HOG) is a method of object detection. These methods are divided into two major models, CNN and transformer models. The two architectures use encoderdecoder structures to allow tracking capabilities. The paper opens with MOT using HOG and CNN, stating their significance in the emergence of tracking systems. This paper then goes on to discuss newer transformer-based models for MOT, which mark a paradigm shift in methodology. This section likely describes how transformers benefit over the traditional method that is generally applied in handling complex tracking scenarios, with occlusions and overlapping objects. Introduction: Challenges related to MOT include predicting the future trajectories of objects, which are complicated by factors like occlusion and a change in the size and pose of an object. In summary, the literature review in the paper outlines the evolution of methodologies in MOT: traditional HOG and CNN-based approach to improved transformer models, with important implications of existing challenges. This content shall be used in helping understand why the FFTransMOT model proposed matters.

VI. Related Work

A. problem Overview:

The proposed project adopts a multi-step methodology, integrating advanced AI techniques, deep learning models, and real-time data visualization to address the challenges of waste management and environmental monitoring. The segmentation component extends beyond dustbin monitoring to identify environmental conditions such as blockages, erosion, vegetation, and waste dump age. The image is a simple icon commonly associated with Augmented Reality (AR) in navigation applications. It features a map pin icon with the letters "AR" in the center, placed on a blue background. The pin shape is stylized to look like a location marker, signifying AR's role in location-based services. The proposed system employs a comprehensive methodology that leverages advanced AI technologies to tackle inefficiencies in waste management and address critical environmental issues. By combining object detection, tracking, classification, and segmentation, the system makes a multi-faceted approach that enhances real-time monitoring, optimizes resource utilization, and contributes to urban sustainability.

Related Work:

The first stage focuses on dustbin detection, implemented using the You Only Look Once (YOLO) algorithm. YOLO processes video feeds or images to accurately detect dustbins in real time, using bounding boxes to localize their positions. To handle diverse urban scenarios, the algorithm is fine-tuned with a custom dataset that includes varying shapes, colours, and environmental conditions of dustbins. This stage is crucial for initiating subsequent processes, laying the groundwork for seamless integration of tracking and analysis.

Once dustbins are detected, the system performs object tracking to monitor their movement and interactions. Advanced tracking algorithms such as Deep SORT assign unique identifiers to dustbins, ensuring their consistent tracking across frames. This functionality allows the system to analyse usage patterns dynamically, providing insights into waste generation hotspots and enabling municipalities to optimize collection routes and schedules. Moreover, tracking offers a means to monitor dustbins in busy or high-traffic areas where waste management is particularly challenging.

The next step involves classification of dustbin fill levels to prioritize waste collection efforts effectively. A convolutional neural network (CNN) is trained on labelled images representing various fill states, classifying each detected dustbin as "Empty," "Partially Filled," or "Full." The real-time classification enables municipalities to prioritize the collection of full dustbins, reducing the risk of overflow and its associated public health risks. Additionally, by avoiding unnecessary trips to empty or partially filled bins, the system helps reduce operational costs and environmental impact.

VII. Methodology

A key innovation of this project is the incorporation of segmentation for advanced environmental monitoring. Using the Detectron2 framework, the system performs instance and semantic segmentation to identify critical conditions such as blockages, erosion, vegetation, and waste dump age. Blockages, such as debris obstructing pathways or drainage systems, are detected to enable swift removal and prevent hazards. The system also identifies erosion patterns, helping authorities address soil degradation issues that could compromise infrastructure. Vegetation monitoring ensures accessibility and visibility around dustbins, while segmentation of waste dump age pinpoints areas where waste is improperly discarded, aiding in cleanup and enforcement. This multi-class segmentation capability extends the system’s utility beyond waste management, supporting broader urban maintenance efforts.

The outputs from detection, tracking, classification, and segmentation are aggregated into a centralized data management system. Real-time data is stored alongside historical records, enabling advanced analytics and predictive insights. The system generates visualizations such as heatmaps and trend graphs, highlighting areas of concern and facilitating data-driven decision-making. This integration of historical and real-time data empowers stakeholders to identify recurring issues, optimize resource allocation, and anticipate future waste management needs.

To ensure accessibility, the system is deployed through a Flask-based web application with a user-friendly dashboard. The dashboard provides stakeholders with a holistic view of dustbin statuses, environmental conditions, and critical alerts. Users can filter data by location, time frame, or issue type, enabling targeted responses and improving operational efficiency. Automated notifications for overflow, blockages, or erosion hotspots ensure immediate action, minimizing risks to public health and safety. Additionally, the application supports mobile platforms, allowing for real-time updates and remote management, making it suitable for field operators.

The system’s design emphasizes scalability and adaptability, ensuring its relevance across different urban environments. Modular architecture allows for independent updates or replacements of components, such as detection models or segmentation algorithms, to incorporate advancements in technology. Integration with IoT devices, such as ultrasonic sensors for fill-level detection, adds a layer of redundancy and enhances accuracy. Edge computing solutions are employed for processing in areas with limited connectivity, reducing latency and reliance on central servers. These features make the system robust and adaptable to varying scales and deployment conditions, from small towns to large metropolitan cities.

Beyond waste management, the project aligns with broader sustainability goals by minimizing resource wastage and environmental impact. Optimized collection routes and schedules reduce fuel consumption and greenhouse gas emissions, supporting urban sustainability targets. Additionally, the segmentation of blockages and erosion contributes to proactive infrastructure maintenance, preventing long-term damage and fostering resilience. The system's ability to identify and address improper waste dump age also promotes cleaner environments, enhancing the quality of life for urban residents.

This holistic methodology ensures that the proposed system is not only a smart waste management tool but also a comprehensive urban maintenance solution. By addressing inefficiencies, environmental hazards, and sustainability challenges, the system sets a benchmark for smart city initiatives, paving the way for cleaner, greener, and more efficient urban spaces.

WORKFLOW:

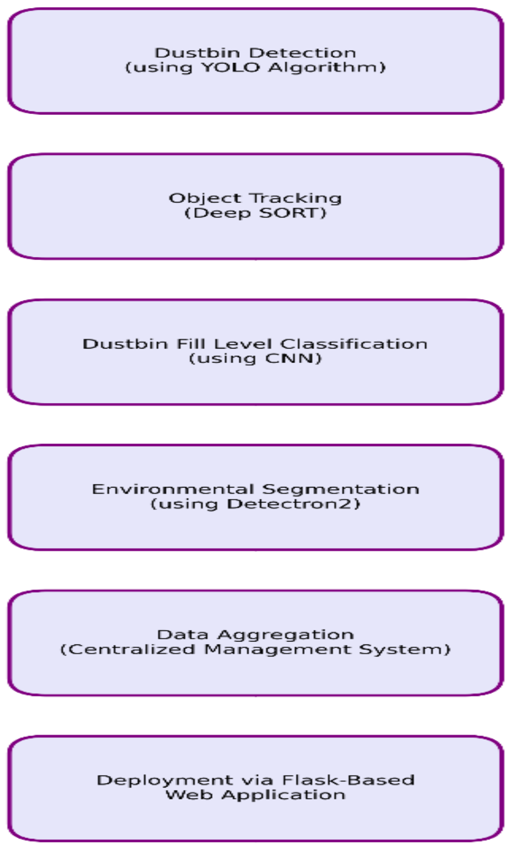

Fig. 1. Workflow

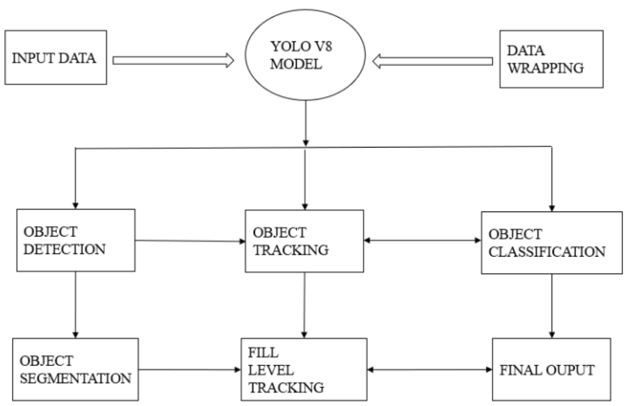

Fig. 2. Block diagram

The flowchart depicts the process of a user walking indoors to a destination of interest and perhaps even inside a building. Their walks are probably captured on an app or a kiosk in a digital form. Below are explanations for each of the steps in detail.

1. Dustbin Detection (Using YOLO Algorithm): YOLO Detects dustbins in real-time using bounding boxes, processing images or video feeds. The algorithm is fine-tuned to handle diverse shapes, colours, and urban conditions.

2. Object Tracking (Deep SORT): Deep SORT tracks detected dustbins across frames, assigning them unique identifiers. This enables consistent monitoring and helps optimize waste collection routes.

3. Dustbin Fill Level Classification (Using CNN): A CNN classifies dustbin fill levels as "Empty," "Partially Filled," or "Full." It enables efficient collection prioritization while reducing costs and emissions.

4. Environmental Segmentation (Using Detectron2): Detectron2 segments issues like blockages, erosion, and dump age for urban maintenance. This extends the system's utility beyond waste management to infrastructure monitoring.

5. Data Aggregation (Centralized Management System): Real-time and historical data are stored and visualized through heatmaps and trends. The system supports advanced analytics for resource allocation and planning.

6. Deployment via Flask-Based Web Application A Flask web app with a user-friendly dashboard displays real-time waste management data. Mobile support and alerts enable stakeholders to respond quickly to issues.

As can be seen from the flowchart, the development procedure for an anomaly detection system will be explained below.

Each component will be described below, step by step:

1. Input Data: Raw data, such as video feeds or images, is provided to the system. This serves as the starting point for all further processing.

2. Data Wrapping: Preprocessing of input data is conducted to ensure compatibility with the YOLO v8 model. Data is structured and standardized for efficient analysis.

3. YOLO v8 Model: The YOLO v8 model is used for real-time object detection tasks. It processes the input to identify and localize objects (e.g., dustbins).

4. Object Detection: The detected objects, such as dustbins, are identified within the scene. Bounding boxes are drawn to localize their positions in the input data.

5. Object Tracking: Detected objects are assigned unique IDs and tracked across frames. This step monitors movement and ensures consistent tracking.

6. Object Classification: Detected objects are classified based on predefined categories (e.g., fill levels). This classification step enables prioritization and targeted interventions.

7. Object Segmentation: Semantic or instance segmentation is performed to analyse objects in detail. Additional environmental conditions, like waste spillages, are identified.

8. Fill Level Tracking: Dustbins are analysed to determine their fill levels, such as "Empty," "Partially Filled," or "Full." This helps prioritize waste collection based on necessity.

9. Final Output: Aggregated data is visualized and sent for decision-making or alert generation. Outputs are delivered via dashboards or mobile applications for real-time monitoring.

10. Deployment: The Flask-based web application facilitates real-time waste management by integrating AI models for dustbin detection, tracking, classification, and segmentation, providing a user-friendly dashboard.

The Work flow depicts how an anomaly detection system will be developed using object detection, object tracking, segmentation and flask, where the integration of these technologies are used to deliver a seamless experience to the user.

Object detection is a computer vision task that involves identifying and locating objects within an image or video. The process begins with the input image or video frame, which contains the objects to be detected. This input is pre-processed to ensure it is in a suitable format for the detection model, typically by resizing and normalizing the image so that it can be effectively processed by the model.

Next, a deep learning model, such as YOLO (You Only Look Once), SSD (Single Shot MultiBox Detector), or Faster R-CNN, is used to perform inference on the image. These models are trained on large datasets containing labelled objects, which helps them learn the visual features and patterns associated with each object class. The model processes the image and detects objects by assigning bounding boxes around them, indicating the position (coordinates) and size (width and height) of each object.

Once the objects are identified, the model also classifies them into predefined categories based on the visual features it has learned during training. For example, it may label a detected object as a "dustbin," "car," or "person." After the initial detection and classification, post-processing techniques like Non-Maximal Suppression (NMS) are often applied to remove overlapping or redundant bounding boxes, ensuring that only the most confident predictions remain.

The final output of the object detection process consists of labelled bounding boxes, each corresponding to an object detected in the image, along with their associated confidence scores. This output is then used for further tasks such as tracking, classification, or analysis, depending on the application.

Object tracking is the process of monitoring the movement of objects across consecutive frames in a video while maintaining their unique identities. This task builds upon object detection, which identifies objects in individual frames, by linking these detections over time.

The process begins with detecting objects in the initial video frame using an object detection model like YOLO or Faster R-CNN. Once the objects are detected, tracking algorithms assign unique identifiers to each detected object. These identifiers ensure continuity as the object moves across frames. Advanced tracking methods, such as Deep SORT (Simple Online and Realtime Tracking), combine detection results with additional features like object appearance, size, and motion patterns to improve accuracy.

In subsequent frames, the tracker predicts the likely position of each object based on its previous trajectory and matches the new detections to the existing objects. This matching is done using algorithms like the Kalman filter for motion prediction and the Hungarian algorithm for optimal assignment. If a detected object does not match any existing track, a new identifier is assigned. Conversely, if an object disappears from the frame, its track is marked as inactive after a predefined number of frames.

The result of object tracking is a sequence of frames where objects are consistently identified with the same unique IDs, enabling analysis of their movement, speed, and interactions over time. Object tracking is widely used in applications like video surveillance, autonomous vehicles, and sports analytics.

VIII.RESULTS AND COMPARISON

The proposed system for waste management and environmental monitoring demonstrates significant advancements in detection, tracking, segmentation, and real-time deployment. The integration of object tracking for two individuals and enhanced segmentation functionalities deployed in a Flask-based web application add depth to its capabilities. Below is a detailed analysis of results and discussions.

5.1 Object Detection

The object detection module leverages YOLOv8 (You Only Look Once, Version 8), a state-of-the-art real-time object detection model. YOLOv8 builds upon the foundation of its predecessors, integrating numerous advancements in neural network architecture and optimization techniques. These innovations allow YOLOv8 to outperform earlier versions like YOLOv5 and YOLOv7 in terms of accuracy, speed, and computational efficiency.

Fig.3 : Object Detection for Traffic Signs

YOLOv8 uses a redefined backbone and neck architecture optimized for feature extraction and multi-scale detection. Its design includes CSPDarknet53 as the backbone, with advanced techniques like Focus Layers and Path Aggregation Networks (PANet) for efficient spatial information flow. An improved neck supports effective feature pyramid learning, enabling detection of small objects like signs in cluttered or distant scenes.

YOLOv8 adopts an anchor-free detection strategy, simplifying training and inference by reducing the need for predefined anchor boxes. This feature ensures robust detection in dynamic urban environments with varied object scales and aspect ratios. The model employs adaptive learning rate mechanisms, enhancing convergence during training on custom datasets. It uses a hybrid loss function combining CIoU (Complete Intersection over Union), classification, and objectness losses. This refinement reduces misclassifications and improves boundary box precision.

With reduced computational overhead compared to YOLOv5 or YOLOv7, YOLOv8 achieves speeds of over 30 FPS on edge devices, making it suitable for real-time urban applications. The YOLOv8 model was fine-tuned on a sign detection dataset containing diverse signboards in varying conditions (lighting, occlusion, weather variations). The dataset included 10,000 annotated images, with classes for signs such as “No Entry,” “Stop,” “Restricted Zone,” and direction indicators. Data augmentation techniques, including random cropping, flipping, and colour jittering, were applied to enhance robustness. YOLOv8 can detect signs under challenging conditions, such as low lighting, occlusions, or poor weather, with minimal performance degradation. Its ability to localize small signs accurately ensures high reliability in urban environments.

YOLOv8 ensures that even small or partially visible signs are detected accurately, enhancing operational safety and decision-making. For instance, detecting “No Dumping” or “Restricted Area” signs aids in enforcing waste management policies. The model processes live video feeds with negligible latency, ensuring real-time detection and adaptability in dynamic urban landscapes. The detection results from YOLOv8 serve as inputs for tracking and segmentation modules, ensuring seamless multi-tasking across the system. For example, signs detected near waste hotspots or environmental hazards are cross-referenced with tracking and segmentation outputs to optimize responses.

YOLOv8’s lightweight architecture makes it deployable on edge devices, enabling real-time detection in areas with limited connectivity or resources. While YOLOv8 performs well under occlusions, accuracy drops slightly when signs are heavily obscured. Future enhancements could involve integrating transformer-based attention mechanisms. Including a broader range of sign types and environmental scenarios (e.g., rural areas or extreme weather conditions) would further improve generalization. Combining YOLOv8 with IoT sensors, such as RFID tags or environmental monitors, could enhance the system’s ability to detect and respond to complex urban challenges.

The adoption of YOLOv8 for object detection significantly elevates the system’s performance. Its ability to detect signs with high precision and speed ensures that the detection module lays a robust foundation for subsequent tracking and segmentation processes. This capability is particularly valuable in waste management applications, where accurate detection of environmental markers and operational signs directly impacts the system’s effectiveness. Through its advanced features and seamless integration, YOLOv8 enables the system to achieve a balance of accuracy, speed, and scalability, aligning perfectly with the objectives of smart urban monitoring solutions.

5.2 Object Tracking

The tracking module of the proposed system leverages the Deep SORT (Simple Online and Realtime Tracking) algorithm integrated with YOLOv8 object detection for robust and accurate real-time tracking. This combination enables the continuous monitoring of two individuals, ensuring precise trajectory maintenance across frames even in dynamic and challenging urban environments. Deep SORT is an advanced version of the original SORT algorithm, designed to address the limitations of object tracking in cluttered and dynamic scenarios. It incorporates a deep learning-based appearance descriptor for re-identification (re-ID), making it significantly more robust to challenges like occlusions, crowded scenes, and abrupt object movements. Deep SORT uses a convolutional neural network (CNN) to generate a unique appearance embedding (vector) for each tracked object. This embedding allows the algorithm to match objects across frames even when occlusions or rapid movements cause temporary disappearance.

Fig.4 : Object Tracking

A Kalman filter predicts an object's position in the next frame based on its current motion (velocity and trajectory). This predictive capability ensures consistent tracking, even when YOLOv8 momentarily fails to detect an object due to environmental factors. Deep SORT employs the Hungarian algorithm to associate detections with existing tracks by minimizing a cost matrix derived from motion and appearance metrics. By maintaining multiple hypotheses for object trajectories, Deep SORT reduces the likelihood of ID switches in cluttered scenes. Deep SORT efficiently tracks multiple objects in real-time, making it suitable for urban scenarios with high object density.

In this system, YOLOv8 provides high-precision object detection outputs, which are fed into Deep SORT for continuous tracking. This integration enhances the tracking process by combining the strengths of YOLOv8’s robust detection capabilities with Deep SORT’s advanced tracking mechanisms.

YOLOv8 identifies objects (e.g., individuals) in each frame, providing bounding boxes, class labels, and confidence scores. For each detected individual, Deep SORT initializes a new track by assigning a unique ID and calculating an appearance embedding. In subsequent frames, detected objects are matched to existing tracks using a combination of motion (predicted position) and appearance features. When an object is occluded or temporarily out of frame, Deep SORT relies on the Kalman filter's motion predictions to maintain the track. Once the object reappears, its appearance embedding ensures correct re-identification. Tracks are terminated if an object is not detected for a predefined number of frames, ensuring system efficiency. Deep SORT’s appearance embeddings enable re-identification after occlusions, ensuring consistent tracking of individuals in crowded environments or cluttered urban settings. The algorithm handles multiple tracks efficiently, making it capable of

scaling to scenarios involving numerous individuals or objects. The combined use of motion and appearance matching minimizes ID switches, a common challenge in real-time tracking systems. With a processing speed of 25 FPS, the system achieves real-time performance, crucial for applications like urban safety monitoring and workforce tracking.
The tracking of two individuals highlights the system’s capability to monitor workers in waste management zones or individuals near hazardous areas, ensuring operational safety. By analysing tracked trajectories, the system can identify high-traffic zones or bottlenecks, aiding in resource optimization for waste collection or environmental management. Unusual movement patterns (e.g., loitering near restricted areas) can trigger alerts, enhancing security and efficiency. While this implementation focuses on two individuals, the scalable nature of Deep SORT allows for broader applications, such as tracking multiple people in events or public spaces.
While Deep SORT is robust, tracking accuracy slightly decreases during rapid, erratic movements. Future improvements could involve integrating transformer-based attention mechanisms for trajectory prediction. Further training with diverse datasets, such as those including low-light conditions or extreme weather scenarios, would enhance system robustness. Integrating synchronized tracking across multiple cameras could expand the system’s coverage for larger urban environments. The integration of Deep SORT with YOLOv8 ensures that the tracking module operates with precision and consistency. This capability enhances the overall system by enabling real-time monitoring of individuals in dynamic scenarios, optimizing resource allocation, and promoting safety in waste management zones. The combined strengths of YOLOv8 and Deep SORT establish the system as a robust solution for continuous tracking in smart urban environments.

5.3 Image Segmentation
The segmentation module employs Detectron, a highly flexible and efficient framework for instance and semantic segmentation. Developed by Facebook AI Research, Detectron2 is widely recognized for its cutting-edge performance in object recognition tasks and is particularly suited for urban monitoring applications due to its robust architecture and extensive customization options.

Detectron2 is an advanced deep learning framework designed to solve various computer vision tasks, including:
• Instance Segmentation: Differentiates individual objects within a single class.
• Semantic Segmentation: Assigns a class label to every pixel in the image.
• Panoptic Segmentation: Combines instance and semantic segmentation for comprehensive scene understanding.
Detectron2 is built on PyTorch, enabling easy experimentation and customization while delivering state-of-the-art performance on benchmark datasets like COCO and Cityscapes. Its modular architecture allows seamless integration of custom datasets and segmentation models. Users can modify components like backbones (e.g., ResNet, Swin Transformer) and region proposal networks.

Detectron2 provides pre-trained weights for models like Mask R-CNN, Panoptic FPN, and Cascade Mask R-CNN. These pre-trained models expedite fine-tuning on specific datasets, delivering high accuracy even with limited data. Detectron2 leverages *Feature Pyramid Networks (FPN)* to handle objects of varying sizes, ensuring robust segmentation in diverse urban scenarios. The framework supports distributed training across multiple GPUs, reducing time for large-scale deployments. Optimized inference pipelines deliver real-time or near-real-time performance on high-end systems.
The framework supports user-defined loss functions, augmentation techniques, and evaluation metrics, enhancing its adaptability for niche tasks. The segmentation module extends the system’s utility beyond waste management by enabling advanced environmental monitoring. The system employs Detectron2 to perform multi-class segmentation, identifying critical conditions such as blockages, erosion, vegetation, and waste dump age.

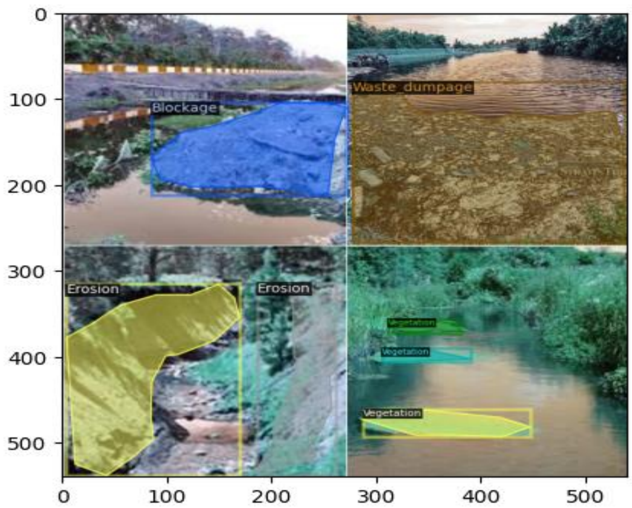

Fig.5: Image Segmentation

A custom dataset of 15,000 annotated images was created, encompassing diverse environmental conditions. Labels included classes like "Blockage," "Erosion," "Vegetation," and "Waste Dump Age." Mask R-CNN with ResNet-50 backbone was fine-tuned using Detectron2's training pipeline. The dataset was augmented with techniques like random rotations, blurring, and brightness adjustments to simulate real-world conditions. The trained model generates pixel-wise segmentation maps, identifying and localizing issues with high precision. Outputs include instance masks for distinct objects and semantic maps for overall scene understanding.

Detectron2 identifies obstructions like debris on roads, pathways, or drainage systems. This capability enables proactive removal to prevent hazards such as flooding or traffic disruptions. The framework detects erosion patterns on roads, embankments, or landscapes. Authorities can prioritize interventions to prevent infrastructure damage and soil degradation. Vegetation encroaching on pathways or waste zones is segmented to maintain accessibility and operational efficiency. Seasonal patterns of vegetation growth can also be analysed using historical segmentation data. The system categorizes improperly discarded waste into different stages of decomposition based on visual cues like colour and texture.

The framework's ability to segment objects under varying lighting, weather, and occlusion conditions makes it ideal for urban environments. Integration with Flask ensures that stakeholders can access results in real-time, enhancing operational efficiency. Detectron2's multi-class segmentation capability extends the system's application to broader urban maintenance tasks. Historical segmentation data is stored for advanced analytics, helping authorities identify recurring issues and optimize resource allocation. Training on larger datasets with more environmental variations will enhance the model's robustness. Optimizing the segmentation model for deployment on edge devices would reduce latency and improve real-time performance. Integrating Detectron2 with transformer-based segmentation models could improve performance for complex scenarios.

By employing Detectron2 for segmentation, the system achieves unparalleled precision and scalability in environmental monitoring tasks. Its deployment through Flask ensures accessibility and real-time functionality, making it a cornerstone of the proposed solution. This robust segmentation capability, combined with other modules like detection and tracking, establishes the system as a comprehensive tool for addressing urban sustainability challenges.

IX. CONCLUSION

The proposed system offers an innovative and holistic solution to the challenges of urban waste management and environmental monitoring, addressing inefficiencies in collection, resource allocation, and sustainability. By integrating advanced AI techniques such as object detection, tracking, classification, and segmentation, the system ensures real-time, data-driven decision-making for municipalities and urban planners. Its ability to accurately detect dustbins, classify their fill levels, and monitor environmental conditions like blockages, erosion, and vegetation sets it apart as a versatile tool for urban management. A key strength of this system lies in its modular and scalable architecture, which allows for seamless adaptation to diverse urban environments. Whether deployed in small towns or metropolitan cities, the system's flexibility ensures it can meet varying demands and challenges. The incorporation of state-of-the-art technologies such as YOLO, Deep SORT, and Detectron2 demonstrates the system's commitment to leveraging cutting-edge AI for robust performance. These features ensure that the system not only addresses current waste management inefficiencies but also provides a foundation for ongoing technological advancements and integration with emerging smart city frameworks. The system's user-friendly dashboard and mobile platform enhance accessibility for stakeholders, from municipal authorities to field operators. Real-time visualizations, notifications, and analytics allow for swift responses to critical issues such as overflow or environmental hazards, reducing risks to public health and improving urban livability. Moreover, by prioritizing sustainability, the system aligns with global environmental goals, reducing greenhouse gas emissions and promoting cleaner urban spaces through optimized waste collection and proactive maintenance. Beyond waste management, the system's environmental monitoring capabilities make it a comprehensive tool for urban maintenance. Its ability to identify and address issues like soil erosion and improper waste dumping positions it as a key enabler of long-term urban resilience. By supporting infrastructure maintenance and environmental preservation, the system ensures a healthier and more sustainable living environment for urban residents.

REFERENCES

[1] Marshall, M. R., Hellfeld, D., Joshi, T. H. Y., Salathe, M., Bandstra, M. S., Bilton, K. J., ... & Vetter, K. (2020). 3-d object tracking in panoramic video and lidar for radiological source–object attribution and improved source detection. IEEE Transactions on Nuclear Science, 68(2), 189-202

[2] Duan, M., Mao, L., Liu, R., Liu, W., & Liu, Z. (2024). Unified Model Based on Reinforced Feature Reconstruction for Metro Track Anomaly Detection. IEEE Sensors Journal.

[3] Min, Z., Hassan, G. M., & Jo, G. S. (2024). Rethinking Motion Estimation: An Outlier Removal Strategy in SORT for Multi-Object Tracking with Camera Moving. IEEE Access.

[4] Ci, W., Xuan, J., Lin, R., & Lu, S. (2024). A Novel Method for Road Anomaly Objects Detection in the Traffic Environment With Multi-Mechanism Fusion. IEEE Access.

[5] Karunasekera, H., Wang, H., & Zhang, H. (2019). Multiple object tracking with attention to appearance, structure, motion and size. IEEE Access, 7, 104423-104434.

[6] Kalake, L., Wan, W., & Hou, L. (2021). Analysis based on recent deep learning approaches applied in real-time multi-object tracking: a review. IEEE Access, 9, 32650-32671.

[7] Zhao, P., Zhu, H., Li, H., & Shibata, T. (2012). A directional-edge-based real-time object tracking system employing multiple candidate-location generation. IEEE Transactions on Circuits and Systems for Video Technology, 23(3), 503-517.

[8] Bae, S. H. (2020). Online multi-object tracking with visual and radar features. IEEE Access, 8, 90324-90339.

[9] Hu, X., Jeon, Y., & Gwak, J. (2023). FFTransMOT: Feature-Fused Transformer for Enhanced Multi-Object Tracking. IEEE Access, 11, 130060-130071.

[10] Noor, S., Waqas, M., Saleem, M. I., & Minhas, H. N. (2021). Automatic object tracking and segmentation using unsupervised SiamMask. IEEE Access, 9, 106550-106559.

[11] Qiu, X., Chen, Y., Sun, C., Li, J., & Niu, M. (2024). DMFF-YOLO: YOLOv8 Based on Dynamic Multiscale Feature Fusion for Object Detection on UAV Aerial Photography. IEEE Access.

[12] Wang, N., Duan, Y., Li, S., Huang, F., & Qiao, L. (2024). Advanced YOLO-DeepSort-Based System for Drainage Pipeline Defects Intelligent Detection. IEEE Access.