AI ENABLED ROBOT FOR DATA COLLECTION IN UNREACHABLE AND EXTREME ENVIRONMENTS

A PROJECT REPORT

Submitted in partial fulfillment for the award of the Degree of

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

(SEMESTER VIII)

2020-2024 Batch

APJ ABDUL KALAM TECHNOLOGICAL UNIVERSITY

By

ALEENA FRANCIS (UNT20CS007)

UNIVERSAL ENGINEERING COLLEGE

VALLIVATTOM, THRISSUR

MAY 2024

DECLARATION

I undersigned hereby declare that the project report "AI ENABLED ROBOT FOR DATA COLLECTION IN UNREACHABLE AND EXTREME ENVIRONMENTS", submitted for the partial fulfillment of the requirements for the award of degree of Bachelor of Technology of the APJ Abdul Kalam Technological University, Kerala is a bonafide work done by us under the supervision of my guide Dr. Nitha C Velayudhan, Professor, UEC. This submission represents my idea in my own words and where ideas or words of others have been included; I have adequately and accurately cited and referenced the original sources. I also declare that I have adhered to ethics of academic honesty and integrity and have not misrepresented or fabricated any data or idea or fact or source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the institute and or the University and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been obtained. This report has not been previously formed the basis for the award of any degree, diploma, or similar title of any other University.

Vallivattom
Date: 07/05/2024

Signature:
Name of Student: ALEENA FRANCIS

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

UNIVERSAL ENGINEERING COLLEGE

VALLIVATTOM, THRISSUR

BONAFIDE CERTIFICATE

This is to certify that the project report titled "**AI ENABLED ROBOT FOR DATA COLLECTION IN UNREACHABLE AND EXTREME ENVIRONMENTS**" is submitted by " **ALEENA FRANCIS (UNT20CS007)** " to the APJ Abdul Kalam Technological University in partial fulfilment of the requirements for the award of the Degree of Bachelor of Technology in Computer Science and Engineering is a bonafide record of the project work carried out by her under our guidance and supervision. This report in any form has not been submitted to any other University or Institute for any purpose.

PROJECT GUIDE	**CO-ORDINATOR**	**HEAD OF THE DEPT.**
Dr. Nitha C Velayudhan	Ms. Najla Nazar	Dr. Nitha C Velayudhan
Professor	Assistant Professor	Professor

Vallivattom
07/05/2024

ACKNOWLEDGMENT

I wish to record my indebtedness and thankfulness to all those who helped me to prepare this report titled **“AI ENABLED ROBOT FOR DATA COLLECTION IN UNREACHABLE AND EXTREME ENVIRONMENTS”** and present it in a satisfactory way.

First and foremost, I thank God Almighty for His providence and for being the guiding light throughout the project.

I feel highly privileged to mention **Dr. Jose K Jacob -Principal, UEC** for his co-operation and help. I deeply and wholeheartedly thank **Dr. Nitha C Velayudhan, HOD, CSE** for her extremely valuable advice and encouragement.

I am especially thankful to my coordinator **Ms. Najla Nazar, Assistant Professor, CSE** for giving me valuable suggestions and critical inputs that helped me in the completion of my project. I would also like to extend our heartfelt thanks to my guide, **Dr. Nitha C Velayudhan, Professor, CSE** for her meticulous guidance and support that helped me in the completion of my project.

I would like to extend my sincere gratitude to all faculties of the Computer Science and Engineering department for the support and suggestions that helped me in the development of my project to what it is now. I thank my parents for the mental support provided during the course of the main project at the times when my energies were the lowest.

ABSTRACT

This paper outlines a new approach to data collection in hazardous or inaccessible environments, introducing an autonomous robot designed for such tasks. The robot utilizes AI technology, specifically human recognition capabilities, integrated into drones. By employing advanced pattern recognition techniques like YOLO (You Only Look Once), the drones achieve real-time human detection with high precision. Equipped with cameras, GPS, and other sensors, these robots aim to revolutionize data collection and analysis in challenging settings. The system combines technologies such as GPS tracking, obstacle avoidance, altitude holding features, and the YOLOv8 algorithm for object detection, promising exceptional performance in various scenarios like environmental monitoring, surveillance, or search and rescue operations. The methodology involves steps like pre-processing video inputs, running the object detection model, post-processing, and periodic re-detection of objects. Testing using the COCO dataset demonstrated the effectiveness of the approach across diverse environmental conditions.

CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI	Artifical Intelligence
CNN	Convolutional Neural Network
YOLO	You Only Look Once
SSMD	Single Shot Multi Box Detector
LiDAR	Light Detection and Ranging
COCO	Company Owned Company Operated
PAFPN	Path Aggregation Feature Pyramid Network
UAV	Unmanned Aerial Vehicle
AP	Average Precision
S-BIRD	Sewer-Blockages Imagery Recognition Dataset
IoU	Intersection over Union
NF-Net	Normalizer-Free Networks
HOG	Histograms of Oriented Gradients
SVM	Support Vector Machines
FPPW	False Positives Per Window
HPWREN	High Performance Wireless Research and Education Network
ResNet	Residual Network
VGG	Visual Geometry Group
R-CNN	Region-based Convolutional Neural Network
CBAM	Convolutional Block Attention Module

SoftNMS	Soft Non-Maximum Suppression
SIFT	Scale-Invariant Feature Transform
DL	Deep Learning
DFWS	Device-Free Wireless Sensing
ML	Machine Learning
WSS	Wireless Sensing System
SOT	Single-Object Tracker
MOT	Multi-Object Tracker
FPS	Frames Per Second
SSH	Self-Supervised Harmonization
LUT	Lookup Tables
SDNet	Scale Decoupled Networks
UAM	Upsampling Aggregation Module
RepVGG	Reparameterized Visual Geometry Group

CHAPTER 1

INTRODUCTION

The " AI Enabled Robot for Data Collection in Unreachable and Extreme Environments" stands as a remarkable fusion of artificial intelligence and robotics, ushering in a new era of technological innovation. Object detection in images and videos refers to the process of using computer vision to identify and locate specific objects in images or videos. This function is important in many applications such as autonomous driving, surveillance, medical and virtual reality. YOLO (You Only Look Once) model has gained popularity in recent years thanks to its integration and accessories. End-to-end optimization makes them a solution for search operations. Unlike traditional object detection methods that handle multiple steps such as localization, elimination, and classification, YOLO's model performs these operations simultaneously in a single neural network, making the search faster and better. Main advantages of YOLO model is characterized by the ability to provide high results during target detection. YOLO provides accurate localization and classification of objects in the image by dividing the input image into grids and predicting the bounding box and classes that will appear in each grid cell.

AI robots collecting data in ultra-poor environments represents a groundbreaking combination of artificial intelligence (AI) and robots and marks a major leap forward in technology innovation. In the rapid development of drone technology, our system has evolved into a solution that seamlessly integrates complex object detection capabilities using the YOLO (Pick One) framework with intuitive motion control capabilities. This includes using the YOLOv8 standard, which has In addition, the application provides instant updates on the detection of human activity, creating positive feedback between the user and the drone. YOLO's perfect combination of object detection capabilities and easy control opens a new era in unmanned aerial vehicle (UAV) applications with the high impact of normal surveillance and security. YOLOv8 powered drones have several important features that strengthen their good work.

First of all, its high accuracy is due to the use of the highest target to detect algorithms that reduce the likelihood of false alarms and negative effects. Additionally, the drone's adaptability to different environments, lighting, and weather conditions ensures consistent performance in a variety of operational situations. Fundamentally, our innovation represents a paradigm shift in detection and control where technologies can be combined to transform the UAS environment. As we embrace the future of quality, efficiency and diversity, we expect practice changes to occur across multiple industries to meet these changes and challenges.

1.1 PURPOSE OF PROJECT

The project's goal is to create an AI-powered robot capable of entering and navigating challenging or dangerous environments where humans can't easily go. Equipped with advanced sensors and AI algorithms, this robot can gather data, analyze it, and transmit crucial information back to operators or systems. By leveraging AI and robotics, the project aims to enable data collection and analysis in places like disaster zones or extreme terrains, ensuring safer exploration and aiding fields like disaster response, environmental monitoring, and industrial applications.

1.2 SCOPE OF PROJECT

The scope of our drone system, seamlessly integrating YOLO-based human presence detection with web app, extends across a diverse range of applications. Primarily, the project finds its niche in the domains of surveillance and security, offering a robust solution for monitoring large areas and critical infrastructure. Beyond security, the drone's capabilities prove invaluable in search and rescue missions, aiding in locating individuals in challenging terrains. Environmental monitoring, particularly in sensitive ecosystems, becomes feasible, contributing to conservation efforts and regulatory compliance. In event security and crowd management, the drone excels at detecting anomalies and ensuring public safety. Industries benefit from the system in inspections, while agriculture sees improved farm management. Additionally, the drone's potential spans research initiatives, educational exploration, and applications in customs, border control, law enforcement, and public safety. As we welcome this innovative fusion of cutting-edge technology, the project heralds a new era in UAV applications, promising enhanced capabilities for real-world challenges.

1.3 PROBLEM STATEMENT

The problem statement for the drone system that detects human presence using YOLO revolves around addressing key challenges and gaps in current surveillance, security, and monitoring systems. The existing methodologies may lack real-time, efficient, and multi-scale detection of human presence, limiting their applicability in various scenarios. Traditional surveillance systems may struggle to provide comprehensive coverage, especially in large or complex environments, resulting in potential security vulnerabilities. Additionally, the lack of seamless control mechanisms poses usability challenges, hindering the rapid deployment and adaptability of drone technology for enhanced aerial presence detection.This project aims to tackle these issues by developing an innovative drone system that leverages the advanced YOLO algorithm for precise human presence detection. The integration of YOLOv8, with its structural enhancements, strives to improve the accuracy and efficiency of the detection process, overcoming limitations posed by conventional surveillance methods. The accompanying web app serves as a central hub, providing users with an intuitive interface for receiving real-time updates on detected human activity. By addressing these challenges, the project seeks to usher in a new era of UAV applications, offering a sophisticated solution for surveillance, security, and various industries grappling with the need for efficient human presence detection and control.

CHAPTER 2

LITERATURE SURVEY

Ning Zhang, Francesco Nex, George Vosselman and Norman Kerle, [1], the research focuses on improving human detection in disaster scenarios using deep learning techniques. Existing models trained on standard datasets like COCO fail to detect disaster victims due to differences in context and appearance. Victims in debris are often only partially visible, and their colors blend with the surrounding debris, making them challenging to detect. To address this, the paper proposes a framework to generate composite images by combining human body parts with debris backgrounds. These synthetic images, crucial for training due to the difficulty in obtaining real victim images, undergo a process using a deep harmonization network to enhance their realism. We select YOLOv5l as the most suitable model, and experiments show that using composite images for training. This approach significantly boosts detection accuracy, demonstrated by a 19.4% improvement in average precision (AP). Additionally, employing these harmonious images further enhances the victim detection model, resulting in a total AP increase of 10.2%. The research contributes by introducing a method to create synthetic training data specifically tailored for detecting human body parts in disaster debris, critical for post-disaster search and rescue operations. The integration of a deep harmonization network significantly enhances the credibility and effectiveness of these synthetic images in training the detection model, offering promise for real-time deployment on unmanned aerial vehicles (UAVs) for autonomous post-disaster search missions.

Ravindra R. Patil, Rajnish Kaur Calay, Mohamad Y. Mustafa, Saniya M. Ansari [2], the research focuses on field of artificial intelligence (AI), computer vision plays a crucial role by simulating human vision to intelligently interpret and recognize visual information. Similar to human perception, this technology depends on integrating extensive datasets and human expertise in a meticulously structured manner, resulting in precise and accurate outcomes. The challenging task of identifying and addressing obstructions within sewer systems, due to their diverse nature and the lack of robust techniques, has spurred innovative research.This research utilizes the ”S-BIRD” dataset, a comprehensive collection of frames depicting sewer blockages, as the foundational training data for a deep neural network model. The YOLOv5 architecture serves as the framework, and the strategic implementation of transfer learning and fine-tuning techniques using the corresponding dataset enhances the model's performance, leading to optimal outcomes. The trained model exhibits exceptional accuracy in detecting sewer blockages, thereby enhancing the reliability and efficiency of the associated robotic framework designed for the proficient removal of diverse blockages.Noteworthy is the achieved mean average precision (mAP) score of 96.30% at a confidence threshold of 0.5, maintaining a consistently high performance level of 79.20% across Intersection over Union (IoU) thresholds ranging from 0.5 to 0.95. This research represents a significant advancement in AI-driven solutions for modern urban sanitation systems. Anticipated contributions include a substantial improvement in the effectiveness of sewer maintenance and a broader application of AI technologies to address challenges within urban infrastructure.

Zhengxin Zhang, [3], the article introduces Drone-YOLO, a set of advanced object detection algorithms specifically tailored for detecting objects in unmanned aerial vehicle (UAV) imagery, which presents distinct challenges such as large image sizes, small-sized objects, dense distribution, overlapping instances, and inadequate lighting. To overcome these hurdles, the Drone-YOLO algorithms, based on the YOLOv8 model, propose several enhancements, particularly focusing on the neck component of the model. They incorporate a three-layer PAFPN structure and a specialized detection head for small-sized objects using large-scale feature maps, significantly improving the algorithm's capability to detect smaller targets in UAV imagery. Moreover, they introduce the sandwich-fusion module into each layer of the neck's up-down branch, combining network and low-level features to provide detailed spatial information across different layers. This fusion utilizes depthwise separable evolution, balancing parameter costs and a broader receptive field. Additionally, the algorithms utilize RepVGG modules in the network backbone, enhancing the learning of multi-scale features and surpassing traditional convolutional layers. Evaluation on the VisDrone2019 dataset demonstrates that the proposed Drone-YOLO (large) outperforms other baseline methods in object detection accuracy, exhibiting substantial improvements in mAP0.5 metrics compared to YOLOv8. Notably, the parameter-efficient Drone-YOLO (tiny) with fewer parameters performs comparably or even better than methods with higher parameter counts on the dataset, validating the efficacy of Drone-YOLO algorithms for object detection in drone imagery.

Peng Zhang, Weimin Lei, Xinlei Zhao, Lijia Dong and Zhaonan Lin, [4], presents the task of crowd counting holds crucial significance in several domains, including video surveillance, accident prediction, public security, and intelligent transportation. However, it encounters several challenges. Firstly, high-density crowd aggregation in public places often leads to severe occlusions, complicating accurate counting. Errors in annotating large-scale datasets also impact training outcomes.

Moreover, in densely populated images, the sizes of human head targets vary widely, making it challenging for a single network to identify both near and far targets concurrently. Current crowd counting methods, typically relying on density plot regression, struggle to differentiate features between distant and near targets, limiting their adaptability to scale changes and impacting detection performance in sparser population areas.To address these issues, a solution is proposed a novel adaptive multi-scale network based on convolutional neural networks (CNN). This network aims to count dense populations while achieving a balance between accuracy, inference speed, and overall performance. To enable the model to distinguish between near and far features, stacked convolution layers are employed to deepen the network's depth. Different receptive fields are allocated based on the distance between targets and the camera, enhancing feature extraction for nearby pedestrians. Depth information is utilized to differentiate between distant and near targets of various scales. Additionally, the original image is divided into four patches for pixel-level adaptive modeling of the population.Furthermore, density normalized average precision (nAP) indicators are introduced to assess spatial positioning accuracy. The proposed NF-Net is validated on challenging benchmarks, including Shanghai Tech Part A and B, UCF_CC_50, and UCF-QNRF datasets. Comparative analysis against state-of-the-art methods demonstrates its superior performance across diverse scenarios. Notably, on the UCF-QNRF dataset, the method effectively resolves complex background interference, showcasing its efficacy in countering intricate challenges in crowd counting tasks.

M.D. Mursalin and Syed Mohammed Shamsul Islam, [5], they introduce delves into utilizing the unique features of human ear shapes for various applications, such as biometric identification, 3D ear reconstruction, gender recognition, genetic studies, and clinical asymmetry analysis. While most ear detection methods rely on 2D images, the paper highlights the limitations of these approaches in handling unconstrained scenarios due to sensitivity to lighting and pose variations. To overcome these limitations, the study proposes leveraging 3D images and explores the advancements in 3D imaging techniques that have accelerated applications in biometrics, robotics, medical diagnosis, and autonomous driving.However, conventional convolutional neural networks (CNNs) face challenges in directly processing irregularly ordered 3D point clouds, prompting the introduction of novel architectures like PointNet and PointNet++ that can analyze features directly on 3D point clouds. These advancements have paved the way for solving various research queries in classification and semantic segmentation tasks within the 3D domain.The research introduces EarNet, a modified version of PointNet++, specifically designed to detect ears directly from 3D point clouds of profile face data. To address pose variations, a rotation augmentation block is included during transfer learning of EarNet. Since labeled 3D point cloud data for ear detection is scarce, the study proposes a novel method for generating synthetic 3D profile face data to train EarNet. Furthermore, a method is devised to create ground-truth labels on real 3D data for quantitative evaluation of EarNet's performance.The contributions of the work encompass the proposal of EarNet, a modified deep learning model, the synthetic generation of 3D profile face data for training, a method for creating ground-truth labels on real 3D data, and comprehensive experiments showcasing state-of-the-art performance on the largest publicly available 3D profile face dataset.

Dalal. N, Triggs. B, [6], introduces a method for human detection utilizing Histograms of Oriented Gradients (HOG) features, aiming to analyze local object appearances and shapes by evaluating well-normalized histograms of image gradient orientations in a dense grid. This approach involves fine-scale gradients, precise orientation binning, relatively coarse spatial binning, and high-quality local contrast normalization in overlapping descriptor blocks, surpassing the performance of Haar wavelet-based detectors for person detection. Describing an existing pedestrian detection system, the paper employs linear Support Vector Machines (SVM) as a foundational classifier and evaluates its performance across various implementation choices. This system demonstrates exceptional accuracy on the MIT pedestrian test set and successfully handles a more challenging dataset containing over 1800 pedestrian images. Moreover, the system's feature set proves effective for other shape-based object classes. The proposed system also utilizes linear SVM as its baseline classifier, showcasing efficiency by processing a 320x240 scale-space image (equivalent to 4000 detection windows) in less than a second. It significantly reduces false positive rates to 89% FPPW (False Positives Per Window), a considerable improvement compared to other detectors. Furthermore, it performs well across a more complex dataset with various pose variations and backgrounds in over 1800 annotated human images. There are plans to develop an advanced detection process utilizing a coarse-to-fine or rejection-chain style detector based on HOG descriptors to further improve the detection process. The paper introduces a new and challenging pedestrian database and outlines future work to optimize the linear SVM detector and integrate motion information using block matching or optical flow fields for continued advancements in the system.

Govil. K, Welch. M.L, Ball. J.T, [7], introduces the surge in larger, more devastating wildfires worldwide, notably in regions like California, the USA, Australia, and South America, is attributed to several factors, including rising temperatures, droughts, increased accumulation of fuel due to aggressive fire suppression methods, dead vegetation, and higher population density near wildlands. To manage these escalating wildfires effectively, early detection becomes crucial, prompting the need for various response strategies that emphasize early notification for evacuation, suppression, and monitoring. Efforts to improve early detection involve a range of sensors, from satellites like GOES, MODIS, and VIIRS to aerial vehicles, mountain-top and ground-level optical cameras. While geosynchronous satellites offer broad coverage, their limited resolution can hinder early fire detection. On the other hand, lower altitude orbiting satellites provide better resolution but take longer to revisit specific areas. Deploying high altitude aircraft or long-duration balloons faces challenges related to coverage, energy consumption, and cost-effectiveness. Optical cameras strategically positioned on fire towers and mountain tops enable rapid smoke detection within minutes of a fire's ignition, especially when combined with optical zoom capabilities. Networks such as HPWREN and Alert Wildfire in California utilize these cameras. Although several commercial products and machine learning models exist for smoke detection, studies suggest that systems like FireWatch outperform them, albeit with a median detection time of about 40 minutes. However, recent advancements showing the ability to detect smoke within minutes with high accuracy indicate promising progress in this domain. Combining ground-level camera systems with satellite approaches appears to be the most effective approach, as they complement each other, bolstering overall detection capabilities.

Loey. M, ElSawy. A, Afify. M, [8], presents the deep learning has significantly transformed computer vision and is now making strides in agriculture, particularly in plant disease detection. Traditional methods relying on conventional machine learning and image processing faced limitations in accuracy and scope. However, the incorporation of deep learning techniques has markedly improved disease detection accuracy and expanded the range of diseases and plant species detectable. Research in this domain heavily leans on deep learning models, utilizing neural networks with multiple layers to grasp intricate patterns and features directly from images. These models automatically extract complex representations from raw data, bolstering disease detection accuracy and resilience. Datasets featuring annotated plant images play a pivotal role in training and validating these deep learning models, ensuring their effectiveness across various plant species and diseases. The choice of deep learning architectures significantly influences detection performance. Convolutional Neural Networks (CNNs), especially models like ResNet, VGG, Inception, EfficientNet, and Transformer-based models, have been adapted and fine-tuned for this purpose, exhibiting diverse levels of performance. Evaluation metrics like accuracy, precision, recall, and F1-score are commonly employed to assess these models accurately. Analyzing research papers in this domain involves examining methodologies, datasets, deep learning model architectures, and achieved performance metrics.Overall, the integration of deep learning into plant disease detection holds promise for advancing agricultural practices, augmenting crop yield, and tackling food security challenges by enabling quicker and more precise identification of plant diseases.

Zhao. Z.Q, Zheng. P, Xu. S.T, Wu. X, [9], delves into the evolution of object detection, emphasizing the transition from traditional methods relying on manually engineered features to the transformative impact of deep learning models. It highlights how deep learning, with its ability to automatically learn hierarchical representations from raw data, has revolutionized object detection, offering advantages in learning semantic, high-level features and alleviating the need for handcrafted features. Discussing various deep learning architectures like Faster R-CNN, YOLO, and SSD, the paper showcases their effectiveness in detecting and localizing objects in images or videos. It emphasizes their scalability and flexibility across different object classes and scenes. Training strategies involving transfer learning, data augmentation, and fine-tuning pretrained models are outlined, showcasing how these techniques enhance the performance and generalization of object detectors, particularly in scenarios with limited annotated data. Furthermore, the paper emphasizes the role of optimization functions and strategies, such as stochastic gradient descent and advanced algorithms like Adam or RMSprop, in efficiently training deep learning models for object detection, ensuring optimized model parameters and faster convergence during training. By exploring specific detection tasks like salient object detection, face detection, and pedestrian detection, the paper underscores their significance in various computer vision applications such as image classification, human behavior analysis, face recognition, and autonomous driving. It acknowledges challenges in object detection due to variations in viewpoints, poses, occlusions, and lighting conditions, prompting ongoing research efforts to address these complexities.Overall, the paper offers a comprehensive review, including experimental analyses, comparing different methods in object detection and related neural network-based learning systems. It aims to provide valuable insights for future directions in these fields, highlighting their mutual influence and advancement in computer vision applications.

Wei. H, Kehtarnavaz. N, [10], introduces a Semi-Supervised Faster Region-Based Convolutional Neural Network, a pioneering approach for detecting individuals and classifying their carried loads in distant video data using high-powered lens cameras. SF-RCNN incorporates computationally efficient image processing steps to identify moving areas signifying a person's presence. This information is fed into a faster RCNN classifier, leveraging transfer learning from ResNet50 for improved frame labeling via semi-supervised learning, enhancing the training process. For load classification, a separate CNN classifier employing GoogleNet transfer learning distinguishes between individuals carrying bundles and those carrying long arms. Despite challenges like low-resolution individuals, heat haze, and camera shake in the video dataset, SF-RCNN outperforms traditional faster RCNN methods, highlighting its significance in the expanding video surveillance market valued at over 35 billion. The research emphasizes the shift towards deep learning approaches in video surveillance, surpassing older methods dependent on manually crafted features. It explores various algorithms in person detection, discussing the rise of CNN-based techniques like RCNN, Fast RCNN, and Faster RCNN in pedestrian monitoring. Challenges posed by far-field video surveillance, such as limited datasets and low image resolution, led to the development of SF-RCNN. This two-stage approach involves a fast-moving area detection method followed by a person detector, and a CNN-based classifier to distinguish between different loads. The method is applied in border monitoring scenarios, aiming to detect illegal crossings or activities, particularly identifying drug bundles and long arms. The study compares SF-RCNN's person detection capabilities with AdaBoost, highlighting substantial advancements.

Kaur R., Singh S [11] introduce the vast field of computer vision, Deep Convolutional Neural Networks (DCNNs) have consistently demonstrated outstanding performance across various applications, including Video Processing, Object Detection, Image Segmentation, Image Classification, Speech Recognition, and Natural Language Processing. Object Detection, among these applications, stands out as both crucial and challenging, finding widespread use in security, military operations, transportation systems, and medical sciences. This review meticulously explores the intricate facets of object detection, recognizing its significance and the evolution of deep learning algorithms dedicated to this task. While acknowledging the notable advancements in object detection models, the review advocates for a nuanced perspective, asserting that conventional methods predating deep learning shouldn't be dismissed. In certain scenarios, traditional approaches leveraging global features prove superior. The review's structure encompasses an overview of object detection, key components like frameworks and convolutional neural networks, commonly used datasets, and discussions on evaluation metrics. It comprehensively addresses challenges, applications, and contemplates future research directions. The concluding section offers a comparative analysis of object detection models, drawing conclusions that encapsulate the field's state at the time. This well-structured examination provides valuable insights into the dynamic landscape of object detection in computer vision.

Patil R.R, Mustafa M.Y, Calay R.K, Ansar S.M.[12] propose the realm of automated and robotic systems, computer vision has emerged as a dependable platform for sewer maintenance and cleaning tasks, particularly with the significant advancements brought about by the ongoing AI revolution. This transformation has notably heightened computer vision's capabilities, especially in discerning issues within underground sewer pipes, such as blockages and damages. The success of AI-based detection models relies crucially on access to substantial, validated, and labeled imagery data, leading to the development of the S-BIRD (Sewer-Blockages Imagery Recognition Dataset) introduced in this paper. The S-BIRD dataset is meticulously crafted to address specific challenges posed by sewer blockages, taking into account parameters like robustness, performance, consistency, and feasibility, particularly in the context of real-time detection tasks. The YOLOX object detection model validates the dataset's utility, illustrating its consistency and viability in effectively identifying sewer blockages. Furthermore, the paper outlines the seamless integration of the S-BIRD dataset into an embedded vision-based robotic system designed for real-time detection and removal of sewer blockages. This integration of computer vision and AI technologies into robotic systems holds immense promise for enhancing the efficiency and precision of sewer maintenance tasks. To emphasize the practical relevance of this work, the paper references the outcomes of an individual survey conducted in Pune, India, a typical mid-size city in a developing country. The survey findings underscore the urgent need for advanced solutions in sewer maintenance, emphasizing the significance of the presented approach in addressing real-world challenges. As cities grapple with the complexities of urban infrastructure, the convergence of computer vision, AI, and robotics in sewer maintenance emerges as a transformative and indispensable tool for creating cleaner, more sustainable urban environments.

Xueshan Gao, Qingfang Zhang, Mingkang Li, Bingqing Lan, Xiaolong Fu, Jingye Li[13] proposes the traversing the intricacies of dynamic obstacle avoidance in man-machine environments presents a formidable challenge for mobile robots dedicated to inspecting hazardous gases. This challenge is particularly heightened in confined spaces like warehouses and laboratories, where the need to account for personal space becomes pivotal. This study conducts a meticulous examination of both human and robot behaviors in such environments, culminating in the development of a comprehensive man-machine social force model. Serving as the foundational framework, this model is instrumental in scrutinizing the optimal speed for robot obstacle avoidance. The investigation delves into the analysis of four distinct man-machine behavior patterns, providing a nuanced understanding that informs the design of an effective robot behavior strategy. Leveraging insights derived from the social force model and behavior patterns, the study introduces an innovative approach the fuzzy-PID trajectory tracking control method coupled with an autonomous obstacle avoidance behavior strategy. This amalgamation is proposed to regulate the mobile robot's speed during the inspection of hazardous gases in relatively confined man-machine dynamic environments. Simulation analysis outcomes underscore the superiority of the proposed controller over the traditional PID control method, showcasing reduced position error (below 0.098 m), angle error (below 0.088 rad), smaller steady-state error, and a shorter convergence time. Crucially, experimental results involving crossing and encountering patterns affirm that the proposed behavior strategy ensures the robot maintains a safe distance from humans while adeptly tracking its trajectory.

Yiting,L, Qingsong Fan, Haisong Huang, Zhenggong Han, Qiang Gu.[14] proposes the tackling the pivotal challenge of Unmanned Aerial Vehicle (UAV) multitarget detection, a matter of great significance for both civil and military applications, this paper grapples with the complexities arising from diverse factors such as varying target sizes, shape changes, occlusion, and dynamic lighting conditions in the realm of drone-based research. The paper introduces an innovative aerial image detection model explicitly crafted to showcase outstanding performance and robustness in the face of these challenges. A noteworthy advancement involves the integration of the Bi-PAN-FPN concept into YOLOv8-s, addressing the specific issues of misdetection and missed detection of small targets in aerial images. This refinement focuses on enhancing the model's neck section, facilitating a more sophisticated feature fusion process that maximizes the utilization of multiscale features while minimizing parameter costs. Additionally, the incorporation of the Ghost-blockV2 structure into the backbone of the benchmark model replaces part of the C2f module, effectively mitigating information loss during long-distance feature transmission and resulting in a substantial reduction in the number of model parameters. Thorough evaluations of the proposed model are conducted on the widely used VisDrone2019 dataset, encompassing comprehensive studies such as ablation studies, contrast analyses, interpretability assessments, and experiments on a self-built dataset. Consistently positive results validate the effectiveness and feasibility of the proposed model, showcasing substantial advantages across various metrics. Beyond pushing the boundaries of UAV multitarget detection, this research provides valuable insights into the real-world application of deep learning. The proposed model holds the potential for enhanced deployment and performance, making a substantial contribution to the progressive evolution of the field.

LiangLiang Zhao, MinLing Zhu, [15] presents the MS-YOLOv7 model stands as a significant breakthrough in Unmanned Aerial Vehicle (UAV) aerial image object detection, serving as an enhancement to the YOLOv7 algorithm. Tailored to surmount challenges prevalent in UAV aerial imagery, such as numerous objects and small-sized entities, its architecture introduces innovative elements. Notably, it incorporates multiple detection heads for simultaneous identification of objects at diverse scales, crucial for handling the varied sizes in UAV images. The integration of a Convolutional Block Attention Module (CBAM) enriches feature extraction across different scales, enhancing the model's comprehension of input images. Addressing challenges in high-density object detection, Swin Transformer units and a novel pyramidal pooling module, SPPFS, are introduced, leveraging their capabilities to capture long-range dependencies and gather multi-scale information, respectively. The model further improves identification of overlapping and occluded objects with Soft Non-Maximum Suppression (SoftNMS) and the Mish activation function. Empirical evaluations on the VisDrone2019 dataset demonstrate the MS-YOLOv7 model's efficacy, revealing substantial performance improvements over other state-of-the-art models. A 6.0% increase in [unspecified metric] and a 4.9% increase in [another unspecified metric] compared to the baseline YOLOv7 algorithm are observed. Ablation experiments confirm the collective contribution of introduced modules to enhanced detection accuracy, and visualizations showcase the model's effectiveness across diverse scenarios. In summary, the MS-YOLOv7 model emerges as a robust and versatile solution adept at addressing complexities in UAV aerial image object detection tasks.

J. Jeyabharathi, Seedha Devi, Bindu Krishnan, Roxanna Samuel, Mohammed Imran Anees, R. Jegadeesan [16] the system introduces an efficient methodology for human ear detection, emphasizing the exploitation of local ear characteristics while tackling challenges such as pose variability, poor contrast, illumination changes, and registration shortages. The approach incorporates parallel image pre-processing techniques to address issues related to noise and illumination. Additionally, it employs the Scale-Invariant Feature Transform (SIFT) process on pre-processed images to minimize pose variability and enhance image validation. The SIFT feature extraction is performed on enhanced images, and a Convolutional Neural Network (CNN) classifier is employed for the overall trial. Evaluation on the IIT Delhi ear database demonstrates that the proposed fusion of techniques significantly enhances recognition accuracy, showcasing its effectiveness in handling diverse challenges associated with human ear detection.

Ronghui Zhang, Xiaojun Jing, Sheng Wu, Chunxiao Jiang, Junsheng Mu, F. Richard Yu [17] article explores the burgeoning field of wireless sensing technologies, specifically focusing on the utilization of wireless signals for passive target wireless sensing in various Internet-of-Things applications. It highlights the potential of deep learning (DL) as a promising technique for device-free wireless sensing (DFWS) and human detection, leveraging recent breakthroughs in machine learning (ML) and artificial intelligence (AI). Despite DL's widespread applications in computer vision, AI games, speech recognition, and automated vehicles, its integration into wireless sensing systems (WSSs) is a relatively novel concept that has received limited attention. The motivation behind the article is to elucidate the mechanisms and potential of DL-aided WSSs for human detection. The survey covers advanced DL architectures suitable for WSSs, reviews traditional ML and DL approaches in human detection using different sensors, and explores the possibility of combining information from heterogeneous sensor types for improved performance in practical human detection systems. The article provides a comprehensive overview of current research in wireless sensing for human detection within WSSs, introduces a detailed structure of DL-based WSS, and identifies open research issues, such as data acquisition, signal calibration, multimodal sensing, simultaneous user identification, activity recognition, multiuser human detection, and DL model generalization, guiding future research directions in this evolving field.

Chung-Cheng Chiu, Wen-Chien Lo [18] addressing the challenging task of object segmentation in real-world applications, this paper presents a novel object detection algorithm that leverages disparity values to overcome issues arising from object overlap and scattered disparities of identical objects. The proposed algorithm employs the connected component algorithm in conjunction with disparity values to establish connections between pixels with similar disparities. Additionally, a recursive object merge technique is applied to consolidate neighboring contours with similar characteristics, facilitating the accurate location and bounding of object contours. The experimental results demonstrate the efficacy of the proposed algorithm in resolving object detection challenges associated with overlapping objects. The algorithm not only successfully addresses the segmentation difficulties but also yields detection results suitable for subsequent object recognition applications, showcasing its potential for practical and comprehensive object analysis in complex real-world scenarios.

Sumaira Manzoor, Kyu-Hyun Sung, Yueyuan Zhang, Ye-Chan An, Tae-Yong Kuc [19] research addresses the critical area of object tracking in real-world scenarios, a subject that has garnered considerable attention due to its broad applications. Despite numerous approaches in both traditional machine learning and modern deep learning for addressing single and multi-object tracking challenges, these tasks remain inherently difficult. The study conducts a comprehensive comparative analysis of eleven object trackers, focusing on determining the most robust single-object tracker (SOT) and multi-object tracker (MOT). The contributions of this work include the use of nine pre-trained tracking algorithms for SOT, such as SiamMask, GOTURN, BOOSTING, MIL, KCF, TLD, MedianFlow, MOSSE, and CSRT. Additionally, the investigation of MOT involves integrating object detection models with object trackers using YOLOv4 with DeepSort and CenterNet with SORT. The researchers develop their own testing videos dataset for experimentation and perform a qualitative analysis based on nine significant factors, including appearance and illumination variations, speed, accuracy, scale, partial and full occlusion, report failure, and fast motion. The experimental results highlight the effectiveness of the SiamMask tracker in overcoming various environmental challenges for SOT, while the YOLOv4 combined with DeepSort tracker shows promising performance for MOT. However, the study also underscores the challenges of handling full occlusion in real-world scenarios and the inherent trade-off between tracking accuracy and speed in the evaluated trackers.

Al-Akhir Nayan, Joyeta Saha, Khan Raqib Mahmud, Abul Kalam Al Azad, Muhammad Golam Kibria [20] it addresses the prevalent issue of image noise, which often arises from factors like insufficient lighting, low-quality cameras, and image compression. While contemporary object recognition techniques and benchmarks primarily focus on images of high quality, there exists a critical need for effective object detection in noisy images, particularly in surveillance and related fields. This paper introduces a novel and cost-effective technique for detecting objects in noisy images, leveraging the advantages of the Single Shot MultiBox Detector (SSD). The proposed approach is subjected to a comprehensive experimental evaluation, comparing its performance with conventional detectors retrained on noisy images. The assessment is conducted using the widely recognized Pascal Visual Object Classes benchmark. Results demonstrate that the suggested technique exhibits satisfactory performance in detecting objects from noisy images when compared to alternative image detection approaches. This method holds promise for application in autonomous industries, offering a solution to object detection challenges arising from inadequate lighting and lower quality images.

Miao Li, Shengjian Mao, Hanzhuo Wang, Zhiguo Shi, Chengwei Zhou [21] the study addresses the escalating concerns related to public safety and personal privacy posed by the rapid proliferation of unmanned aerial vehicles (UAVs). Vision sensors are commonly employed to detect intruding drones due to the accessibility and intuitiveness of video data. However, identifying and positioning long-distance intrusion drones becomes challenging due to their small pixel area and weak morphological features. While existing efforts have primarily focused on enhancing drone detection accuracy, this work aims to concurrently improve detection speed and reduce network training time. The proposed drone detector prioritizes high accuracy, short training time, and fast inference speed. The Centernet architecture is chosen for its light-head, single-stage, and anchor-free design, facilitating rapid inference. The framework of Centernet is enhanced through a separation strategy for identification and positioning features. Introducing a novel Gaussian elliptic coding method with a corresponding loss function helps encode more training samples efficiently, simulating the effect of increased batch size to boost learning rates and accelerate training. Considering the complexity of the background during flight, a background filter is employed to mitigate the impact of intricate backgrounds. The proposed model is trained and evaluated on the Drone-vs-Bird dataset, achieving a mean Average Precision (mAP) of 83.7% and a Frames Per Second (FPS) rate of 29.6, demonstrating its effectiveness in addressing the challenges associated with

UAV detection.

Yifan Jiang, He Zhang, Jianming Zhang, Yilin Wang, Zhe Lin, Kalyan Sunkavalli, Simon Chen, Sohrab Amirghodsi, [22] a novel approach called the Self-Supervised Harmonization framework (SSH) for image harmonization, which aims to enhance the quality of image compositing by aligning the appearance factors like color tone, brightness, and contrast between foreground and background images. Traditional methods for image harmonization require large annotated datasets that involve intricate professional retouching. The SSH framework, in contrast, introduces a self-supervised learning approach that can be trained using unaltered, "free" natural images. The framework redefines the image harmonization problem by adopting a representation fusion perspective, separately processing foreground and background examples to address issues related to background occlusion. This design enables a dual data augmentation method, generating diverse triplets [foreground, background, pseudo ground truth] through image cropping with perturbations using 3D color lookup tables (LUTs). Additionally, the authors create a real-world harmonization dataset curated by expert users for comprehensive evaluation and benchmarking. The results demonstrate that the proposed self-supervised method outperforms previous state-of-the-art approaches in terms of reference metrics, visual quality, and user studies. The code and dataset for SSHarmonization are made publicly available for further exploration and application.

Meroua Belmir, Wafa Difallah, Abdelkader Ghazli [23] the significant impact of plant diseases on agricultural production, emphasizing the need for accurate classification and early detection to mitigate the adverse effects on crop variety and yield. The proposed solution involves the development of a deep convolutional neural network (CNN) model designed for the classification of plant diseases. The model is trained and evaluated using the PlantVillage dataset, comprising images representing 14 different healthy and diseased crop leaves, categorized into 38 distinct classes. The experimental results showcase the effectiveness of the CNN model, achieving an impressive training accuracy of 98.01% and a test accuracy of 94.33%. These high accuracies indicate the model's robust capability in accurately classifying and detecting various plant diseases at an early stage. By leveraging deep learning techniques and utilizing a diverse dataset, this CNN model stands out as a promising tool for farmers and agricultural practitioners to proactively manage and control plant diseases, ultimately contributing to enhanced crop quality and quantity.

Fuyan Lin, Xin Zheng, Qiang Wu [24] focuses on enhancing the effectiveness of deep learning-based object detection methods, particularly in the context of UAV (Unmanned Aerial Vehicle) inspections where automation of line inspection is crucial. The challenge addressed is related to aerial view scenes captured by drones flying at high altitudes, resulting in small proportions of objects in the images. Traditional YoloV3 networks may struggle to achieve satisfactory detection results for small objects due to limited information in the 8x downsampling feature map. To address this limitation, the authors modify the YoloV3 network based on the LaSOT dataset. The modification involves adjusting anchor values and introducing a 4x downsampling prediction layer to enhance the detection of small objects. The improved YoloV3 network demonstrates advancements in convergence ability and detection accuracy when compared to the original YoloV3 network. By tailoring the architecture to better handle the challenges of UAV inspection scenarios, this modified YoloV3 model exhibits improved performance in detecting small objects in aerial images, contributing to the automation and efficiency of line inspection tasks.

Tianwen Peng, Guozhu Liu, Hongtao Liang [25] proposes algorithm addresses the challenge of multi-scale object detection by introducing the Scale Decoupled Networks (SDNets) and Upsampling Aggregation Module (UAM) to enhance precision across different scales. SDNets employs three parallel detection branches dedicated to small, medium, and large-scale objects, aiming to mitigate conflicts between features at different scales. By incorporating convolutions with varying receptive fields, the algorithm extracts features of diverse scales, thereby improving the network's capacity to express them accurately. The Upsampling Aggregation Module (UAM) introduces a pathway for upsampling feature fusion during downsampling, contributing to more effective feature fusion. Experimental results on the VOC dataset demonstrate the algorithm's efficacy, achieving a detection precision of 83.2%, surpassing YOLOX by 2.0% and YOLOv7 by 0.3%. Notably, the algorithm significantly enhances detection precision for small, medium, and large-scale objects by 3.7%, 2.7%, and 2.1%, respectively. This approach showcases improved performance in multi-scale object detection, making it a promising advancement in the field.

CHAPTER 3

FEASIBILITY

3.1 FEASIBILITY STUDY

A project feasibility study is a comprehensive report that examines in detail the five frames of analysis of a given project. It also takes into consideration its four Ps, its risks and POVs, and its constraints (calendar, costs, and norms of quality). The goal is to determine whether the project should go ahead, be redesigned, or else abandoned all together. A feasibility study is an analysis used in measuring the ability and likelihood to complete a project successfully including all relevant factors. It must account for factors that affect it such as economic, technological, legal and scheduling factors. A feasibility study helps identify logistical problems, and nearly all business-related problems and their solutions. Feasibility studies can also lead to the development of marketing strategies that convince investors or a bank that investing in the business is a wise choice.

3.2 TYPES OF FEASIBILITY STUDY

The feasibility of the system is based on the technical, operational, economic and schedule feasibilities. A feasibility study is an analysis that takes all of a project's relevant factors into account including economic, technical, legal, and scheduling considerations to ascertain the likelihood of completing the project successfully. Project managers use feasibility studies to discern the pros and cons of undertaking a project before they invest a lot of time and money into it. Feasibility studies also can provide a company's management with crucial information that could prevent the company fromn entering blindly into risky businesses.

3.2.1 TECHNICAL FEASIBILITY

Technical feasibility study involves leveraging existing technologies like AI, drones, and advanced algorithms such as YOLO for real-time human detection in hazardous environments. The integration of hardware components like cameras, GPS, and sensors suggests a robust system capable of capturing and processing data effectively. Validation through testing with the COCO dataset underscores the system's ability to perform in diverse environmental conditions. However, challenges such as environmental factors and scalability may require further consideration.

3.2.2 OPERATIONAL FEASIBILITY

This measures how well your product/system will be able to solve. It is used for identifying the impotance of certain problem in project and how it is to be solved. It also measures how solution of the problems will work for any project .It analyses the behaviour of the proposed system and whether the proposed system is easier than the existing system for the users of the system.Our system is a comprehensive methodology for data collection in hazardous or inaccessible environments using autonomous robots. By integrating AI technology and advanced algorithms into drones, the system demonstrates a practical approach to real-time human detection and data analysis. Moreover, the deployment of drones equipped with cameras, GPS, and sensors facilitates efficient data gathering while minimizing risks to human operators. The system's versatility, with applications ranging from environmental monitoring to search and rescue operations, highlights its potential for widespread adoption. However, ensuring seamless integration with existing operational frameworks and addressing potential logistical challenges, such as maintenance and personnel training, will be essential for successful implementation. Nonetheless, the project's operational feasibility is promising, offering a viable solution for enhancing data collection capabilities in challenging settings.

3.2.3 ECONOMIC FEASIBILITY

It is used to determine the financials resources of the project. It measures all costs incurred in the development of the new system. It is called cost benefit analysis because it determines the total cost for the development of new system and benefits derived from the new system. The economic feasibility of the project appears promising. While initial investment costs for developing and deploying the autonomous drone system may be significant, the potential long-term benefits justify the expenditure. By revolutionizing data collection in hazardous or inaccessible environments, the system can lead to cost savings and efficiency gains in various sectors, such as environmental monitoring, surveillance, and search and rescue operations. Additionally, the scalability of the technology suggests the possibility of widespread adoption, further enhancing its economic viability. However, factors such as ongoing maintenance costs, regulatory compliance, and market demand should be carefully considered to ensure the project's economic sustainability. Overall, with careful planning and strategic implementation, the project presents a compelling case for economic feasibility and potential return on investment.

3.2.4 SCHEDULE FEASIBILITY

Schedule feasibility is defined as the likelihood of a project being completed within its scheduled time frame. If the project has a high likelihood of completion by the desired due date then schedule feasibility is considered to be high. We studied about the requirements of our project before 2 months of project initiation. After discussing various problems about the existing system we ended with the exact requirements of the new system and after then we designed the system. We made various design model like data flow diagram, block diagram etc.

CHAPTER 4

HARDWARE AND SOFTWARE REQUIREMENTS

4.1 HARDWARE REQUIREMENTS

4.1.1 SERVER REQUIREMENTS

- Operating System : Windows 10/Windows 11
- Processor : Intel processors i5
- Main Memory : 8 GB or above
- Internet Connection
- High rpm Motors, fly controller, Radio controller, Battery
- Web Camera
- Sensors

4.2 SOFTWARE REQUIREMENTS

4.2.1 YOLOV8

YOLOv8, short for ”You Only Look Once version 8,” is an advanced object detection algorithm used in computer vision tasks. It builds upon the YOLO (You Only Look Once) architecture, which is known for its real-time object detection capabilities. YOLOv8 enhances the accuracy and efficiency of object detection by introducing improvements in network architecture, training techniques, and optimization strategies. It employs a single neural network to simultaneously predict bounding boxes and class probabilities for multiple objects within an image. YOLOv8 is widely used in various applications, including surveillance, autonomous vehicles, and robotics, where real-time object detection is crucial.

4.2.2 PYTHON

Python is a high-level programming language renowned for its simplicity, versatility, and readability. Its concise syntax and extensive libraries make it ideal for a wide range of applications, from web development and data analysis to artificial intelligence and automation. Python's interpreted nature allows for rapid development and prototyping, while its cross-platform compatibility ensures seamless deployment across various operating systems. With a vibrant community and robust ecosystem, Python continues to be a popular choice for both beginners and experienced developers alike, driving innovation and powering countless projects worldwide.

4.2.3 PYTHON FLASK

Python Flask is a lightweight and powerful web framework used for building web applications and APIs in Python. It is known for its simplicity, flexibility, and ease of use, making it a popular choice among developers for projects of all sizes. Flask provides essential features for web development, such as routing, request handling, and template rendering, while also allowing for easy integration with other Python libraries and frameworks. Its minimalistic design encourages modular and scalable application development, making it suitable for a wide range of use cases, from simple websites to complex web applications. With its extensive documentation and active community support, Flask empowers developers to create robust and efficient web solutions efficiently.

4.2.4 RASPBERRY PI OS

Raspberry Pi OS, formerly known as Raspbian, is the official operating system for Raspberry Pi single-board computers. It is a Debian-based Linux distribution tailored specifically for the Raspberry Pi hardware architecture. Raspberry Pi OS provides a lightweight and optimized environment for running applications on Raspberry Pi devices, making it ideal for various projects and applications, including programming, web browsing, multimedia, and home automation. It comes with a range of pre-installed software packages, including programming tools, productivity software, and development environments, making it suitable for both beginners and experienced users. Raspberry Pi OS is regularly updated and maintained by the Raspberry Pi Foundation, ensuring compatibility and stability with the latest Raspberry Pi hardware and software developments.

4.2.5 DRONEKIT

DroneKit is an open-source SDK (Software Development Kit) that provides a Python API for controlling and communicating with drones, particularly those running on autopilot software like ArduPilot. It simplifies the process of developing custom applications for drone operations, offering features such as flight control, mission planning, telemetry monitoring, and event handling. With DroneKit, developers can create innovative drone applications for various industries and use cases, leveraging its high-level Python API to interact with drones and perform tasks such as waypoint navigation, geofencing, and real-time data analysis.

4.2.6 HTML

HTML, or HyperText Markup Language, is the standard language used for creating web pages and web applications. It provides a structured way to format content on the web by using a series of tags and elements. HTML documents consist of a hierarchical structure of elements, each representing different parts of a webpage, such as headings, paragraphs, images, links, forms, and more.HTML is complemented by other technologies like CSS (Cascading Style Sheets) for styling and layout, and JavaScript for interactivity and dynamic behavior. Together, these technologies form the backbone of the World Wide Web, enabling the creation of rich and interactive web experiences.

4.2.7 CSS

CSS, or Cascading Style Sheets, is a style sheet language used for describing the presentation of HTML documents. It provides a way to control the layout, formatting, and appearance of web pages, including elements such as text, colors, fonts, spacing, and positioning.CSS enables web developers to create visually appealing and consistent designs across multiple web pages and devices. It allows for flexibility and modularity in styling, making it easier to maintain and update styles across a website. With CSS, developers can achieve a wide range of design effects, from simple color changes to complex layouts and animations, enhancing the user experience and usability of web applications.

CHAPTER 5

IMPLEMENTATION

5.1 SYSTEM ARCHITECTURE

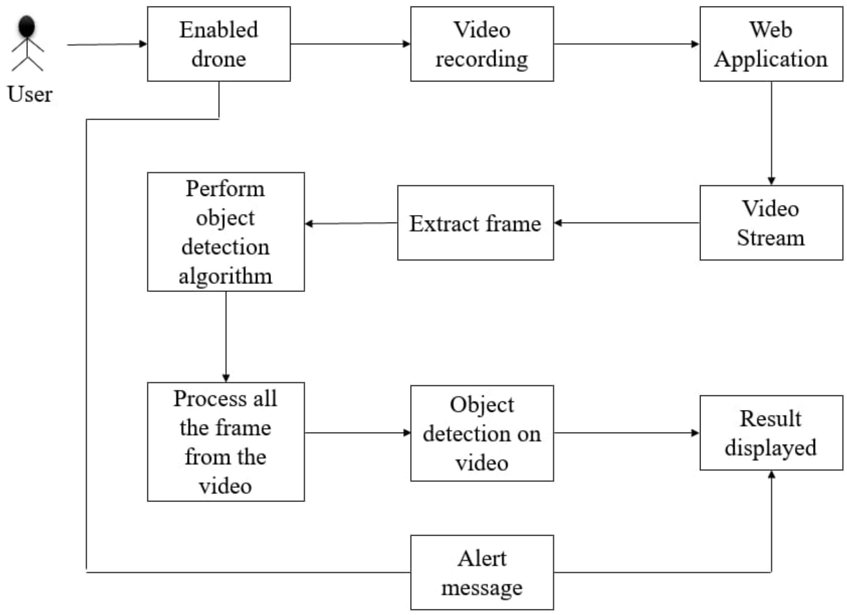


Fig. 5.1: System Architecture.

The project's architecture is designed to seamlessly integrate various components, facilitating efficient interaction and communication between the user, the drone, and the web application. At its core, the user interface serves as the primary means for users to engage with the system, enabling them to initiate drone operations, such as video recording, and receive alerts. This interface acts as a gateway through which users can access and interact with the system's functionalities. Concurrently, the drone plays a pivotal role in capturing video footage of the environment, utilizing its onboard camera to record live streams that are transmitted to the web application in real-time. These video streams are then processed within the web application, where frames are extracted for further analysis. Leveraging advanced object detection techniques, such as YOLOv8, the system identifies objects of interest within each frame, generating bounding boxes and classifications in real-time. These detected objects are subsequently visualized on the user interface, providing users with immediate insights into the environment's dynamics. Moreover, the drone is equipped with alerting capabilities, enabling it to autonomously send alert messages to the user via the web application when specific objects or events are detected. This proactive approach empowers users to respond swiftly to potential threats or anomalies, enhancing situational awareness and facilitating timely interventions. In essence, the project's architecture orchestrates a seamless interplay between user interaction, drone operations, video processing, object detection, and alerting mechanisms, culminating in a comprehensive system for real-time monitoring and response in hazardous or inaccessible environments.

5.2 DESIGN

A system must be designed based on our requirements and a detailed analysis of the system. This particular phase is known as system design. It is the most crucial phase in the development of a system. First, a logical design of the system is already determined. The logical design of a system pertains to an abstract representation of the data flows, inputs, and outputs of the system.

The design process for human detection with a drone starts by analyzing requirements. Then, a system architecture is devised, detailing components and interactions. Components are carefully chosen, including the drone, camera, and controlling system. The human detection algorithm, like YOLOv8, is integrated for analyzing captured data. Rigorous testing ensures the system meets requirements, and a user interface is crafted for displaying detection results. This iterative process ensures the system efficiently detects humans and provides real-time feedback to users.

5.2.1 BLOCK DIAGRAM

A Block diagram is a diagram of the system in which the principal parts or functions are represented by blocks connected by lines that show the relationship of the blocks, they are heavily used in engineering hardware designs, electronics designs, and software designs. Block diagrams are typically used for higher-level, less detailed descriptions that are intended to clarify overall concepts without concern for the details of the information

In the block diagram of AI-enabled drone (Fig 5.2), the process begins with the controlling device, which interfaces with the drone, issuing control signals to direct its movements and actions. Equipped with a camera, the drone then autonomously or as per instructed by the controlling device captures videos and images of its surroundings. These visual data are subsequently transmitted back to the controlling system, likely through a wireless connection. Upon receiving the captured footage, the system employs the YOLOv8 human detection algorithm to process the images and videos, identifying and locating human figures within the scenes. YOLOv8, known for its real-time object detection capabilities, analyzes the data based on predefined patterns and features to accurately detect humans. Once the human detection process is complete, the system presents the results to the user. This presentation could involve displaying a live video feed with bounding boxes around detected individuals or still images highlighting the identified humans, providing the user with real-time information about human presence in the drone's vicinity. Overall, this system integrates drone technology, computer vision, and machine learning to facilitate efficient and effective human detection in various environments.

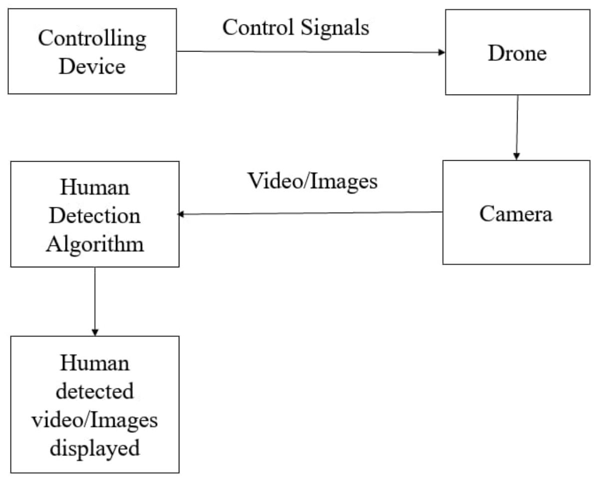


Fig. 5.2: Block diagram.

5.2.2 USE CASE DIAGRAM

The use case diagram is to demonstrate the different ways that a user might interact with a system. A use case diagram at its simplest is a representation of a user's interaction with the system that shows the relationship between the user and the different use cases in which the user is involved. A use case diagram can identify the different types of users of a system and the different use cases and will often be accompanied by other types of diagrams as well.

In a use case diagram of AI-enabled drone (Fig.3) involving drone operations, there are two primary processes: the user's interactions and the system's functionalities. The user process encompasses controlling the drone's movements, receiving notifications, and being alerted to critical events. Users can actively direct the drone, receive status updates, and respond to alerts about detected hazards or targets. On the other hand, the system process involves scanning the environment, acquiring visual data, detecting target areas or objects, and alerting the user based on its analyses. The system autonomously gathers information through sensors, processes images and photos, identifies noteworthy elements in the surroundings, and notifies users of significant findings. This delineation in the use case diagram illustrates the seamless collaboration between user inputs and system capabilities, ensuring efficient and informed drone operations.

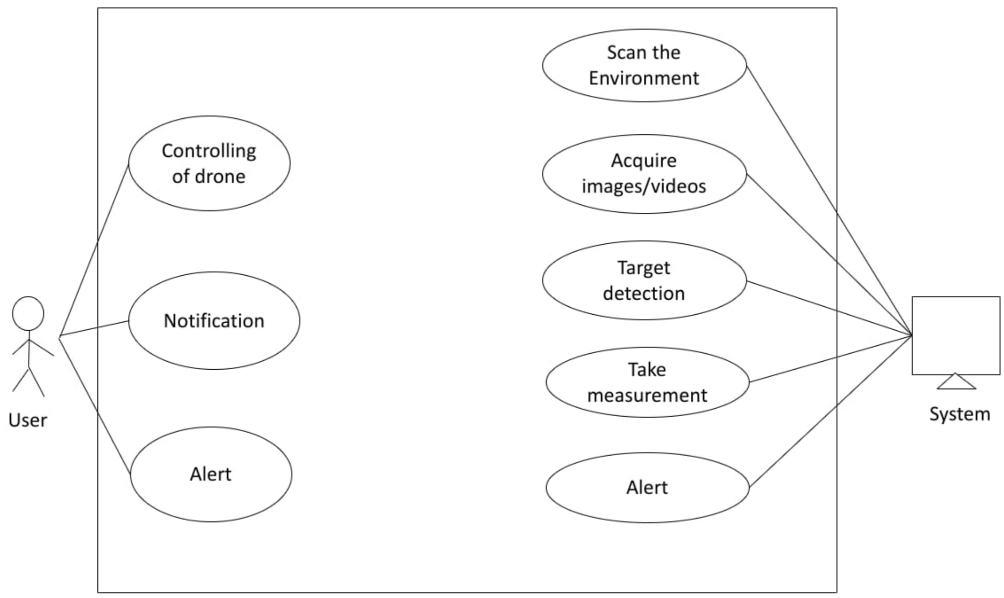


Fig. 5.3: Use Case diagram.

5.2.3 ACTIVITY DIAGRAM

Activity diagram is another important diagram in UML to describe the dynamic aspects of the system. Activity diagram is basically a flowchart to represent the flow from one activity to another activity. The activity can be described as an operation of the system. The control flow is drawn from one operation to another. This flow can be sequential, branched, or concurrent. Activity diagrams deal with all types of flow control by using different elements such as fork, join, etc. The basic purposes of the activity diagrams are similar to the other four diagrams. It captures the dynamic behavior of the system. Other four diagrams are used to show the message flow from one object to another but the activity diagram is used to show the message flow from one activity to another. Activity is a particular operation of the system. Activity diagrams are not only used for visualizing the dynamic nature of a system but they are also used to construct the executable system by using forward and reverse engineering techniques. The only missing thing in the activity diagram is the message part. It does not show any message flow from one activity.

In the activity diagram of AI-enabled drone (Fig.4) outlines a comprehensive process for a system designed to monitor and analyze a specified target area. Initially, the system selects the target area and initializes itself for operation. It then proceeds to record both images and videos within this area, storing them within the system for further analysis. The captured images undergo background extraction, isolating relevant objects, while the system also detects and tracks humans within the footage. Data related to these detections is maintained within the system, ensuring a continuous recording process without interrupting the ongoing analysis. This iterative process of capturing, processing, detecting, and storing data ensures the system's continuous monitoring and analysis capabilities within the defined target area, making it suitable for applications such as surveillance or automated monitoring systems.

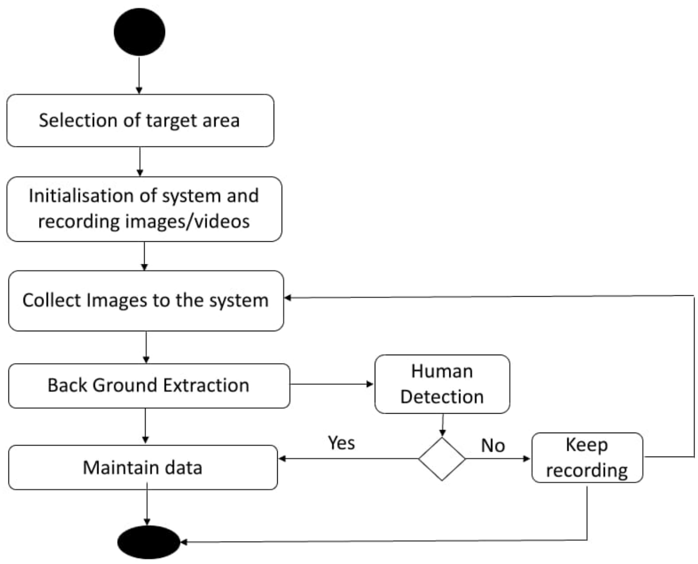


Fig. 5.4: Activity diagram.

5.2.4 DATA FLOW DIAGRAM

A Data Flow Diagram (DFD) maps out the flow of information for any process or system. A Data Flow Diagram shows the way information flows through a process or system. It includes data inputs and outputs, data stores, and the various sub-processes the data moves through. DFDs are built using standardized symbols and notation to describe various entities and their relationships. It uses defined symbols like rectangles, circles, and arrows, plus short text labels, to show data inputs, outputs, storage points, and the routes between each destination. Data flowcharts can range from simple, even hand-drawn process overviews, to in-depth, multi-level DFDs that dig progressively deeper into how the data is handled. They can be used to analyze an existing system or model a new one.

DFD Level 0

It is also known as a context diagram. It's designed to be an abstract view, showing the system as a single process with its relationship to external entities. It represents the entire system as a single bubble with input and output data indicated by incoming/outgoing arrows. Here the system is viewed as a single process, with its relationship to external entities.

At the top level of the Data Flow Diagram (DFD) level 0 for the AI-enabled drone system, the process begins with the drone utilizing its onboard cameras to capture videos and images of its surroundings. Users interact with the drone's camera functionality through the application interface, which includes options for enabling or disabling the camera and configuring various camera settings such as resolution and frame rate. The captured videos and images are then forwarded to the YOLOv8 human detection algorithm for analysis. While YOLOv8 undertakes the task of detecting humans within the visual data using advanced object detection capabilities, users can monitor the process in real-time through the application interface, receiving immediate feedback on the progress and results of the analysis. The system incorporates robust error handling mechanisms to address any potential issues that may arise during data capture, transmission, or analysis, ensuring uninterrupted operation and reliability. Furthermore, the integration of the drone's navigation systems with the human detection process allows for adjustments to flight path or behavior based on detected human presence, enhancing safety and efficiency. This comprehensive system integration seamlessly merges drone technology, user interface design, and sophisticated human detection algorithms to facilitate efficient and reliable human detection across diverse environments.

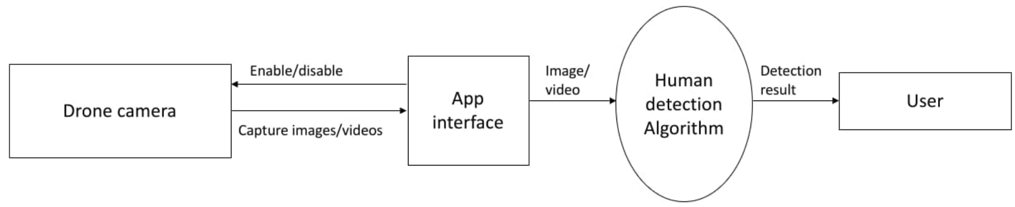


Fig. 5.5: DFD Level 0.

DFD Level 1

The progression from the Level 0 Data Flow Diagram (DFD) involves a detailed breakdown into Level 1 DFD, where fundamental modules within the system are illustrated along with the data flow between these modules. At the Level 1 DFD, essential processes and sources of information are explicitly outlined, offering a more nuanced depiction compared to the higher-level Context Level Diagram. This refined diagram provides a comprehensive and detailed perspective on the system's inner workings, emphasizing the main functions performed by the system as it delves into its sub-processes.

In the Level 1 Data Flow Diagram (DFD) of the AI-enabled drone(Fig.5), the drone's camera captures images and videos, which are then transmitted to the application interface. Within the application, three primary functions are executed. Firstly, the human detection algorithm operates to detect humans or recognize objects within the captured data, forwarding the results back to the application for further processing. Secondly, upon detection of targets, the application generates alert notifications, notifying the user of any identified objects or humans. Lastly, the application facilitates user interaction through the drone control unit, allowing users to send commands to both the drone and the application via keyboards. This comprehensive system architecture ensures efficient image and video processing, target detection, alert generation, and user interaction, enhancing the overall functionality and usability of the drone system.

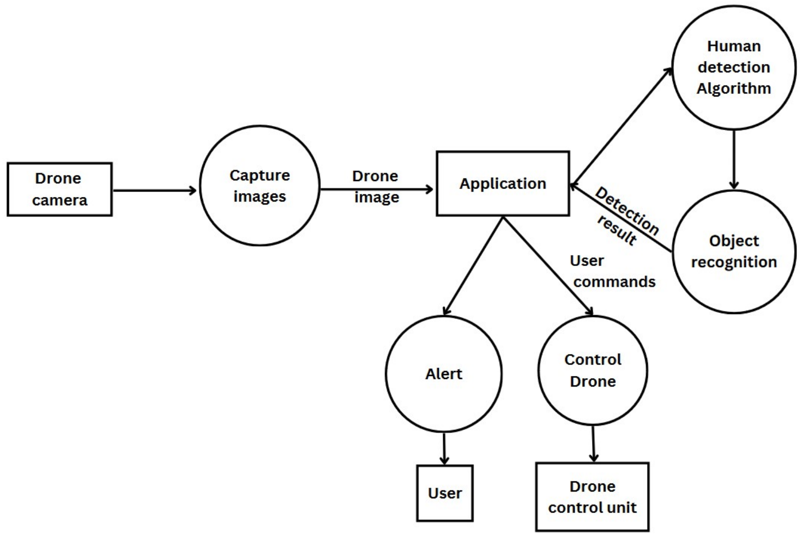


Fig. 5.6: DFD Level 1.

CHAPTER 6

TESTING

6.1 SOFTWARE TESTING

Software testing is a critical process in the development life cycle aimed at identifying defects, errors, or bugs in a software application to ensure its quality and reliability. It involves executing the software under controlled conditions and assessing its behavior against expected outcomes. There are several types of software testing, including unit testing, which tests individual components or units of code; integration testing, which verifies the interaction between integrated components; and system testing, which evaluates the overall functionality of the entire system.

Additionally, software testing includes non-functional testing, such as performance testing to assess the system's responsiveness and scalability, usability testing to evaluate the user interface and experience, and security testing to identify vulnerabilities and ensure data protection. Overall, software testing plays a crucial role in mitigating risks, improving software quality, and delivering a reliable and satisfactory user experience.

6.1.1 SOFTWARE UNIT TESTING

In the software unit testing for an AI-enabled drone tailored for data collection in extreme and unreachable environments, individual components and functionalities of the software are thoroughly examined. This involves isolating and testing each unit of the software, such as algorithms for navigation, data collection, and communication, to ensure they perform as intended under challenging conditions. Unit tests verify the accuracy and reliability of these components, helping to identify and rectify any potential issues early in the development process. By conducting comprehensive unit testing, developers can enhance the overall reliability and effectiveness of the AI-enabled drone software, ensuring optimal performance in extreme environments.Additionally, the software must be tested to ensure seamless communication and coordination between the drone's components, including sensors, cameras, and communication systems, to effectively collect and transmit data back to the user. The software unit testing is essential to ensure the AI-enabled drone can reliably perform its data collection tasks in even the most demanding environments.

6.1.2 INTEGRATION TESTING

Integration Testing is the level of software testing in which individual software modules are combined and tested as a group. It occurs after unit testing and before system testing. Integration testing takes as its input modules that have been unit tested, groups them in larger aggregates, applies test defined in an integration test plan to those aggregates, and delivers as its output the integrated system ready for system testing.

Integration testing for an AI-enabled drone in extreme and unreachable environments involves examining how its components collaborate under adverse conditions. It verifies the cohesion between navigation systems, data collection tools, and communication modules. Testing assesses the drone's ability to operate seamlessly in challenging environments, ensuring reliable data collection and transmission. By simulating extreme conditions, integration testing validates the drone's performance and robustness, crucial for successful missions in such environments.

6.1.3 SYSTEM INTEGRATION TESTING

System testing of software or hardware is testing conducted on a complete, integrated system to evaluate the system's compliance with its specified requirements. System testing falls within the scope of black-box testing and as such, should require no knowledge of the inner design of the coder or logic. System testing takes as its input an integrated software system that has passed the integration testing phase. System testing is usually considered appropriate for assessing nonfunctional system requirements.

System integration testing for an AI-enabled drone tailored for data collection in extreme and unreachable environments scrutinizes the interaction and functionality of the entire system. This testing phase evaluates how all components, including hardware and software, collaborate harmoniously to achieve the intended objectives. It validates the seamless coordination between navigation systems, data collection mechanisms, AI algorithms, and communication modules under adverse conditions. By simulating real-world scenarios and extreme environments, system integration testing ensures the reliability, accuracy, and effectiveness of the AI-enabled drone system in fulfilling its data collection mission in challenging settings.

6.1.4 SYSTEM VALIDATION TESTING

System Validation is a set of actions used to check the compliance of any element (a system element, a system, a document, a service, a task, a system requirement, etc.) with its purpose and functions. These actions are planned and carried out throughout the life cycle of the system.

System validation testing for an AI-enabled drone aimed at data collection in extreme and unreachable environments verifies whether the entire system meets the specified requirements and performs as expected in real-world scenarios. This testing phase involves assessing the drone's ability to navigate, collect data, and communicate effectively in challenging conditions such as harsh weather or remote locations. It validates the accuracy and reliability of the AI algorithms utilized for data analysis and decision-making, ensuring they deliver actionable insights. Additionally, system validation testing evaluates the drone's overall performance, including its endurance, durability, and adaptability to extreme environments. By confirming the system's capability to operate reliably in adverse conditions, system validation testing assures stakeholders that the AI-enabled drone is fit for its intended purpose of data collection in unreachable and extreme environments.

CHAPTER 7

METHOD AND METHODOLOGY

A project can be brought to a successful end in various ways. But the best and most popular project management methodologies, methods, and frameworks are always changing. New concepts appear all the time. An entire string of methods, tools, and techniques lies behind all successful projects. However, project management methods, methodologies, and frameworks are not just for project managers. The entire project team must understand their usage, purpose, and basic terms. This will ensure that the whole process will go smoothly regardless of your choice. Remember that no project or team is the same. A methodology or framework that worked for someone else might not be the right one for you. That's why it's best to test how can use them for the own projects.

7.1 METHOD

Method defines as a procedure or process for attaining an object: such as a systematic procedure, technique, or mode of inquiry employed by or proper to a particular discipline or systematic plan followed in presenting material for instruction. In other words, a method refers to a single action, tool, technique, process, or way of doing something.

Developing an AI-enabled drone system for human detection in inaccessible and extreme environments using YOLOv8 involves several key steps. Initially, you need to compile a diverse dataset containing images or videos depicting humans in various extreme conditions, encompassing different weather and lighting situations, terrains, and distances. This dataset serves as the foundation for training the YOLOv8 model, which excels in real-time object detection due to its ability to predict bounding boxes and class probabilities directly from full images in one pass.

Once the dataset is annotated to indicate human presence in each frame, the YOLOv8 model is trained using this annotated data. The trained model is then integrated into the software stack of AI-enabled drones. This integration enables drones to conduct real-time human detection while navigating through unreachable or hazardous environments. However, ensuring optimal performance requires suitable onboard hardware capable of efficiently running the YOLOv8 model.

Efforts are directed towards optimizing the inference process to achieve real-time performance on the drone's hardware, as well as developing flight planning and navigation algorithms to ensure effective coverage of target areas while avoiding obstacles. An alert and response system is also implemented to react appropriately to detected humans, whether it involves notifying operators, recording video footage, or autonomously tracking or following individuals.

Extensive testing and evaluation of the system in both simulated and real-world scenarios are crucial for assessing its performance, reliability, and effectiveness in extreme environments. Feedback gathered from these deployments informs iterative improvements to address any limitations or challenges encountered in the field. Through this iterative process of data collection, model training, integration, testing, and refinement, an AI-enabled drone system leveraging YOLOv8 can be developed to effectively detect humans in inaccessible and extreme environments, enhancing situational awareness and response capabilities.

7.2 METHODOLOGY

A methodology is essentially a set of guiding principles and processes for managing a project. It is a collection of methods, practices, processes, techniques, procedures, and rules. In project management, methodologies are specific, strict, and usually contain a series of steps and activities for each phase of the project's life cycle. The choice of methodology defines how the work and communication is happened. They are defined approaches that show us exactly what steps to take next, the motivation behind each step, and how a project stage should be performed. It comprises the theoretical analysis of the body of methods and principles associated with a branch of knowledge. A methodology offers a theoretical perspective for understanding which method, set of methods, or best practices can be applied.

7.2.1 OBJECT DETECTION

Object detection using YOLOv8, an advanced variant of the You Only Look Once (YOLO) algorithm, encompasses a sophisticated process aimed at accurately identifying objects within images or video frames. Initially, the algorithm partitions the input images or frames into a grid, typically with dimensions like 7x7 or 13x13, to facilitate localized predictions. For each grid cell, YOLOv8 predicts multiple bounding boxes, each characterized by coordinates representing the center, width, height, and a confidence score indicating the likelihood of containing an object. Concurrently, the algorithm estimates class probabilities for various object categories within each bounding box. Following this prediction phase, YOLOv8 applies non-maximum suppression to refine the detections, eliminating duplicate or overlapping bounding boxes while retaining the most confident predictions. This meticulous post-processing ensures precise object localization and classification. Ultimately, the detected objects are visualized with bounding boxes and class labels, showcasing YOLOv8's efficacy in real-time object detection tasks. With its ability to handle diverse object scales and aspect ratios, YOLOv8 finds extensive utility across domains such as surveillance, autonomous driving, and industrial automation, where rapid and accurate object detection is paramount.

7.2.2 IMAGE PROCESSING

The utilization of image processing technology within the autonomous robot-'s design is a crucial component that significantly contributes to its overall functionality. Image processing serves as a sophisticated mechanism to analyze images captured by the robot's sensors, primarily cameras. These sensors capture visual data from the robot's surroundings, providing a continuous stream of images that encapsulate the environment in which the robot operates. The first step in this process involves preprocessing the images, where various techniques may be applied to enhance image quality, reduce noise, or adjust brightness and contrast, ensuring that the subsequent analysis is performed on high-quality data.

The core of image processing lies in its ability to extract meaningful information from these images. Object recognition, a key task facilitated by image processing, involves identifying and categorizing objects within the visual field. Advanced algorithms and techniques are employed to detect patterns, shapes, and features that distinguish one object from another. This capability is invaluable in scenarios where the robot needs to navigate, interact with its surroundings, or perform specific tasks based on identified objects.

Image processing technology further extends its utility in enhancing the robot understanding of the environment. By extracting relevant features and information from the visual data, the robot gains a nuanced perception of its surroundings. This understanding is critical for decision-making processes, allowing the robot to navigate through complex terrains, avoid obstacles, and adapt to dynamic environments. Additionally, image processing aids in the interpretation of scenes, enabling the robot to recognize landmarks or navigate towards specific points of interest.

In summary, the incorporation of image processing technology within the autonomous robot's architecture is pivotal for its effective functioning. It transforms raw visual data into actionable insights, allowing the robot to recognize and interpret its environment. This capability is fundamental for tasks such as object recognition, navigation, and decision-making, making image processing an indispensable tool for enhancing the overall autonomy and versatility of the robot in complex and dynamic scenarios.

7.2.3 GPS TRACKING TECHNOLOGY

The incorporation of GPS tracking technology in the design of the autonomous robot marks a significant advancement in its navigation and localization capabilities. GPS, or Global Positioning System, provides a reliable and accurate means of determining the robot's geographical coordinates in real-time. This technology relies on a network of satellites orbiting the Earth, and the robot's GPS receiver interprets signals from these satellites to triangulate its exact position. The utilization of GPS ensures precise tracking of the robot's location, allowing it to establish its coordinates with a high degree of accuracy.

With GPS technology integrated into the robot's system, the autonomous vehicle gains the capability to navigate through its environment with a clear understanding of its position relative to the Earth's surface. This is particularly crucial for tasks that require the robot to move within predefined areas or follow specific routes. The availability of accurate location data contributes to efficient path planning and navigation, enabling the robot to traverse complex terrains, avoid obstacles, and reach designated destinations.

Moreover, the GPS tracking technology serves as a foundational element for mapping and monitoring activities. The robot's movements and exploration of the environment can be systematically recorded and mapped using the precise location data provided by GPS. This mapping functionality not only aids in real-time monitoring of the robot's activities but also facilitates post-mission analysis and planning for future tasks. It enhances the overall autonomy of the robot, allowing it to adapt to different environments and effectively carry out its designated missions.

In conclusion, the incorporation of GPS tracking technology is instrumental in enhancing the autonomous robot's navigational capabilities. It ensures accurate and real-time location data, enabling the robot to navigate with precision, plan efficient routes, and contribute to mapping and monitoring tasks. This integration empowers the autonomous robot to operate effectively in diverse environments and perform tasks that require spatial awareness and precise location information.

7.2.4 ALERT SYSTEM

The implementation of a notification system within the autonomous robot's design represents a crucial aspect aimed at bolstering safety and responsiveness, especially in hazardous environments. The primary purpose of this system is to promptly alert human operators or supervisors in the event of critical situations, ensuring swift and informed decision-making. The notification system is integrated as a proactive measure to address unforeseen challenges or emergency scenarios that the autonomous robot may encounter during its operations.

In the context of hazardous environments, the notification system serves as a vital safety feature. It is designed to detect and respond to various critical situations, such as potential collisions, equipment malfunctions, or deviations from planned paths. By promptly notifying human operators, the system allows for immediate intervention, enabling operators to take corrective actions or provide guidance to the autonomous robot in real-time. This proactive approach significantly enhances the overall safety and reliability of the system, reducing the risk of accidents or damage in challenging operational settings.

The responsiveness of the notification system is a key factor in its effectiveness. Upon detecting a critical situation, the system initiates immediate communication with operators through predefined alert mechanisms. This could include visual indicators, auditory alarms, or even real-time notifications delivered to operators' devices. The goal is to minimize response times, ensuring that operators are quickly aware of potential issues and can take timely actions to address them.

In hazardous environments where conditions may change rapidly, the notification system's role becomes paramount. By providing timely alerts, the system not only enhances safety but also contributes to the adaptability and resilience of the autonomous robot in dynamic and unpredictable settings. Operators can receive critical information promptly, enabling them to make informed decisions and maintain control over the robot's operations, even in challenging conditions.

In summary, the implementation of a notification system is a strategic and safety-oriented feature within the autonomous robot's architecture. It acts as a vigilant sentinel, promptly notifying operators of critical situations, thereby enhancing the safety, responsiveness, and overall reliability of the system, especially in hazardous and dynamic operational environments.

CHAPTER 8

RESULT AND DISCUSSION

8.1 DATASET & EXPERIMENTS ENVIRONMENTS

The COCO (Common Objects in Context) dataset stands as a cornerstone in the realm of computer vision, providing researchers and developers with a rich and diverse collection of images annotated with precise object labels and segmentation masks. Spanning a myriad of real-world scenes captured in varying contexts, the COCO dataset encapsulates the complexity of everyday environments, offering a comprehensive representation of common objects encountered in the world. From bustling urban streets to serene natural landscapes, the dataset encompasses a wide spectrum of scenarios, each brimming with a multitude of object categories. These categories range from commonplace entities such as people, animals, and vehicles to more specific items like household objects, electronics, and sports equipment. Furthermore, the annotations within the COCO dataset provide invaluable insights into object localization and segmentation, with meticulously outlined bounding boxes and pixel-level masks delineating the spatial extent of each object instance. Beyond mere object detection, the dataset also serves as a foundation for exploring advanced computer vision tasks, including instance segmentation and image captioning, enabling researchers to delve into the nuances of scene understanding and contextual comprehension. As a result, the COCO dataset has emerged as a gold standard benchmark in the field, driving innovation and fostering advancements in object recognition algorithms, neural network architectures, and semantic understanding models. Its widespread adoption and continuous expansion underscore its significance as a pivotal resource in shaping the future landscape of computer vision research and applications.

In the experiment, we used Ubuntu 20.04 as the operating system with Python 3.10, PyTorch 1.16.0, and Cuda 11.6 as the desktop computational software environment. The experiment utilized NVIDIA 3080ti graphics cards as hardware. The implementation code of the neural network based on the Ultralytics 8.0.105 version. The hyperparameters used during the training, testing, and validation of the experiment remained consistent. The training epoch was set at 300, and the images inputted into the network were rescaled to 640×640.

8.2 EXPERIMENT METRICS

The experiments evaluate proposed methods in terms of detection performance and model parameter size. The experiment metrics include precision (P), recall (R)and F1 score. Precision (P) is the proportion of correctly predicted targets among all detected targets. Where TP represents the correct prediction targets, and FP represents the incorrect prediction targets.

$$P = \frac{TP}{(TP+FP)} \tag{8.1}$$

Recall (R) is the proportion of correct detected targets among all existing targets. Where FN represents targets that exist but have not been correctly detected.

$$R = \frac{TP}{(TP+FN)} \tag{8.2}$$

The F1 score is the harmonic mean of recall and precision, which provides a balanced measure of the model's performance. The formula for the F1 score is

$$F1SCORE = 2*((P*r)/(P+R)) \tag{8.3}$$

Monitoring the box loss and class loss separately helped to identify areas where the model may need improvement, such as in localizing objects or correctly classifying them. Similarly, monitoring the val loss helped to identify when the model is over- fitting to the training data and may need regularization techniques like dropout or weight decay. mAP@0.5 and mAP@0.95 are in the IoU threshold used for evalu- ation. mAP@0.5 measures the average precision across IoU thresholds from 0.5 to 1.0, while mAP@0.95 measures the average precision across IoU thresholds from 0.95 to 1.0.

8.3 VISUALISATION

The visualization of the project's results involves displaying the detected objects and their classifications in real-time on the user interface. This visualization provides users with immediate insights into the environment's dynamics and enables proactive response to potential threats or anomalies.

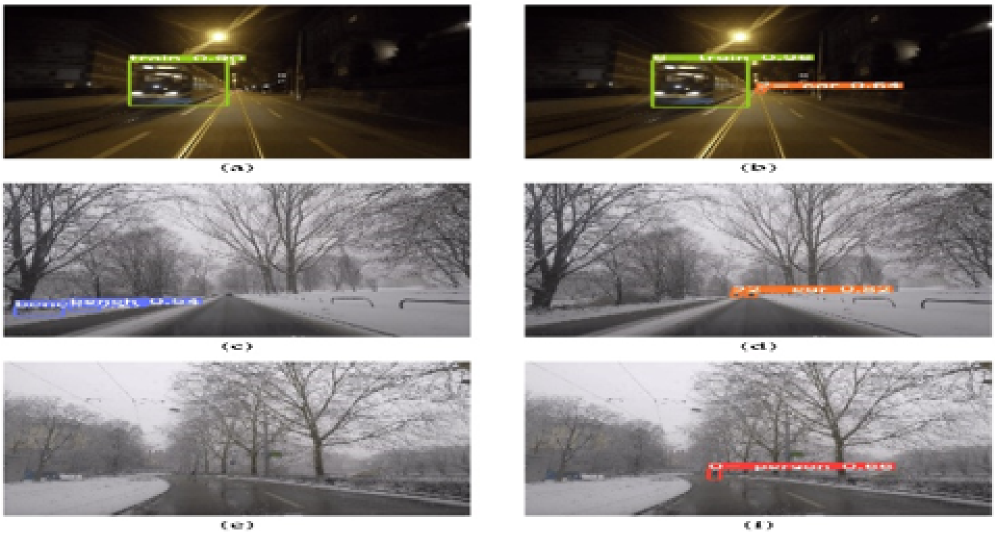


Fig. 8.1: Detection result on different environments.

Additionally, each detected object is accompanied by its corresponding classification or label, indicating the type of object identified (e.g., person, car, or animal). This classification information is displayed alongside or within the bounding boxes, enhancing users' understanding of the detected objects and facilitating rapid decision-making.

CHAPTER 9

FUTURE SCOPE

AI enabled robots offer the advantages of increased efficiency, accuracy, and safety in performing tasks that are either too risky, labor-intensive, or time-consuming for humans to undertake alone. Their ability to adapt to different environments and learn from their experiences makes them versatile tools for a wide range of applications. In search and rescue the AI enabled robots can be crucial in search and rescue operations, especially in areas where it's too dangerous for humans to venture, like collapsed buildings or hazardous terrains. These robots equipped with sensors and AI can navigate through debris, detect signs of life, and relay information back to rescue teams. During natural disasters such as earthquakes, floods, or hurricanes, AI-enabled robots can assist in various tasks like assessing damage, clearing debris, delivering supplies, or even providing medical assistance. Their ability to operate autonomously or under remote control makes them invaluable assets in chaotic and dangerous environments. Detecting Cracks in Structures in infrastructure management, AI-powered robots can be deployed for regular inspections of critical structures like dams, bridges, and pipelines. They can use various sensors such as cameras, lidar, or ultrasonic devices to detect cracks, corrosion, or other signs of deterioration, helping prevent catastrophic failures and ensuring structural integrity. Military and Defense the AI-enabled robots have significant potential in military and defense applications. They can be used for tasks such as reconnaissance, surveillance, bomb disposal, and perimeter security. Their ability to operate autonomously or in collaboration with human forces enhances situational awareness and reduces risks to soldiers' lives.In agriculture, AI-powered robots can revolutionize farming practices by performing tasks such as planting, weeding, monitoring crop health, and harvesting. These robots can use AI algorithms to analyze data from various sources like drones, satellites, and ground sensors to optimize crop management practices, increase yields, and reduce resource consumption.

CHAPTER 10

CONCLUSION

An innovative autonomous robot has been designed and implemented to collect essential data in hazardous or inaccessible environments. The robot is engineered to navigate and collect data from locations too dangerous or remote for human exploration, unlocking unprecedented possibilities for scientific research and exploration. The AI powered drone, capable of precise human identification through a user-friendly mobile control interface, leverages advanced deep learning approaches and architectures like YOLO for remarkable accuracy in realtime human detection tasks. The robot is equipped with a diverse array of sensors, including cameras, and uses advanced image processing and GPS tracking technologies for thorough data collection and processing. The integration of a notification system for critical situations enhances the robot's responsiveness in challenging environments. This technological amalgamation holds the promise of revolutionizing data collection in previously considered unreachable environments, opening new avenues for scientific discovery, resource assessment, and environmental monitoring. The proposed autonomous robot is a catalyst for transformative advances in scientific exploration, resource evaluation, and environmental surveillance in remote and dangerous locations.

This technological amalgamation holds the promise of revolutionizing data collection in previously considered unreachable environments, opening up new avenues for scientific discovery, resource assessment, and environmental monitoring. The proposed autonomous robot emerges as a catalyst for trans-formative advancements in scientific exploration, resource evaluation, and environmental surveillance in remote and dangerous locations. By pushing the boundaries of what is achievable, this technology expands our horizons of knowledge and discovery, marking a pivotal step towards unlocking the full potential of data collection in challenging and unexplored terrains. The comprehensive and sophisticated capabilities of the autonomous robot position it as a key player in reshaping the landscape of data collection methodologies, with far-reaching implications for the advancement of scientific understanding and exploration in challenging terrains.

REFERENCES

[1] Ning Zhang, Francesco Nex, George Vosselman, NOrman kerle, "Training a disaster victim detection network for UAC search and rescue using harmonious composite images". 2022.

[2] Ravindra. R, Patil, Rajnish Kaur Calay, Mohammead. Y, Mustafa, Saniya. M, Ansari, "AI-Driven High-Precision Model for Blockage Detection in Urban Wastewater Systems". 2023.

[3] Zhengxin Zhang, "Article Drone-YOLO: An Efficient Neural Network Method for Target Detection in Drone Images". 2023.

[4] Peng Zhang, Weimin Lei, Xinlei Zhao, Lijia Dong and Zhaonan Lin, "An Adaptive Multi-Scale Network Based on Depth Information for Crowd Counting". 2023.

[5] M.D. Mursalin and Syed Mohammed Shamsul Islam, "Deep Learning for 3D Ear Detection: A Complete Pipeline From Data Generation to Segmentation". 2021.

[6] Dalal. N, Triggs. B, "Histograms of oriented gradients for human detection", In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, 20–25 June 2005.

[7] Govil. K, Welch. M.L, Ball. J.T, Pennypacker. C.R, "Preliminary results from a wildfire detection system using deep learning on remote camera images". Remote Sens. 2020.

[8] Loey. M, ElSawy. A, Afify. M, "Deep learning in plant diseases detection for agricultural crops", A survey. Int. J. Serv. Sci. 2020.

[9] Zhao. Z.Q, Zheng. P, Xu. S.T, Wu. X, "Object detection with deep learning", A review. IEEE Trans. Neural Netw. Learn. Syst. 2019.

[10] Wei. H, Kehtarnavaz. N, "Semi-supervised faster RCNN-based person detection and load classification for far field video surveillance", Mach. Learn. Knowl. Extr. 2019.

[11] Kaur. R, Singh. S, " A comprehensive review of object detection with deep learning." Digit. Signal Process. 2022.

[12] Patil. R.R, Mustafa. M.Y, Calay. R.K, Ansari. S.M, " S-BIRD: A Novel Critical Multi-Class Imagery Dataset for Sewer Monitoring and Maintenance Systems." Sensors 2023.

[13] Xueshan Gao, Qingfang Zhang, Mingkang Li, Bingqing Lan, Xiaolong Fu, Jingye Li, " Mobile Robot Combination Autonomous Behavior Strategy to Inspect Hazardous Gases in Relatively Narrow Man–Machine Environment. ". 2022.

[14] Yiting. L, Qingsong Fan, Haisong Huang, Zhenggong Han, Qiang Gu, " A Modified YOLOv8 Detection Network for UAV Aerial Image Recognition. ". 2022.

[15] LiangLiang Zhao, MinLing Zhu, " YOLOv7 Based on multi-scale for object dtection on UVA Aerial Photography. ". 2023.

[16] J. Jeyabharathi, Seedha Devi, Bindu Krishnan, Roxanna Samuel, Mohammed Imran Anees, R. Jegadeesan, "Human Ear Identification System Using Shape and structural feature based on SIFT and ANN Classifier. ".2022.

[17] Ronghui Zhang, Xiaojun Jing, Sheng Wu, Chunxiao Jiang, Junsheng Mu, F. Richard Yu, "Device-Free Wireless Sensing for Human Detection: The Deep Learning Perspective. ".2021.

[18] Chung-Cheng Chiu, Wen-Chien Lo, "An Object Detection Algorithm with Disparity Values. ".2020.

[19] Sumaira Manzoor, Kyu-Hyun Sung, Yueyuan Zhang, Ye-Chan An, Tae-Yong Kuc,"Qualitative Analysis of Single Object and Multi Object Tracking Models. ".2022.

[20] Al-Akhir Nayan, Joyeta Saha, Khan Raqib Mahmud, Abul Kalam Al Azad, Muhammad Golam Kibria,"Detection of Objects from Noisy Images. ".2020.

[21] Miao Li, Shengjian Mao, Hanzhuo Wang, Zhiguo Shi, Chengwei Zhou,"A Training-time Friendly Network for Real-time Drone Detection. ".2023.

[22] Yifan Jiang, He Zhang, Jianming Zhang, Yilin Wang, Zhe Lin, Kalyan Sunkavalli, Simon Chen, Sohrab Amirghodsi, Sarah K,"SSH: A Self-Supervised Framework for Image Harmonization. ".2021.

[23] Meroua Belmir, Wafa Difallah, Abdelkader Ghazli,"Plant Leaf Disease Prediction and Classification Using Deep Learning. ".2023.

[24] Fuyan Lin, Xin Zheng, Qiang Wu,"Small object detection in aerial view based on improved YoloV3 neural network. ".2020.

[25] Tianwen Peng, Guozhu Liu, Hongtao Liang,"Multi-Scale Object Detection Algorithm Based on Decoupled Networks and Upsampling Aggregation.". 2023.

APPENDIX-1

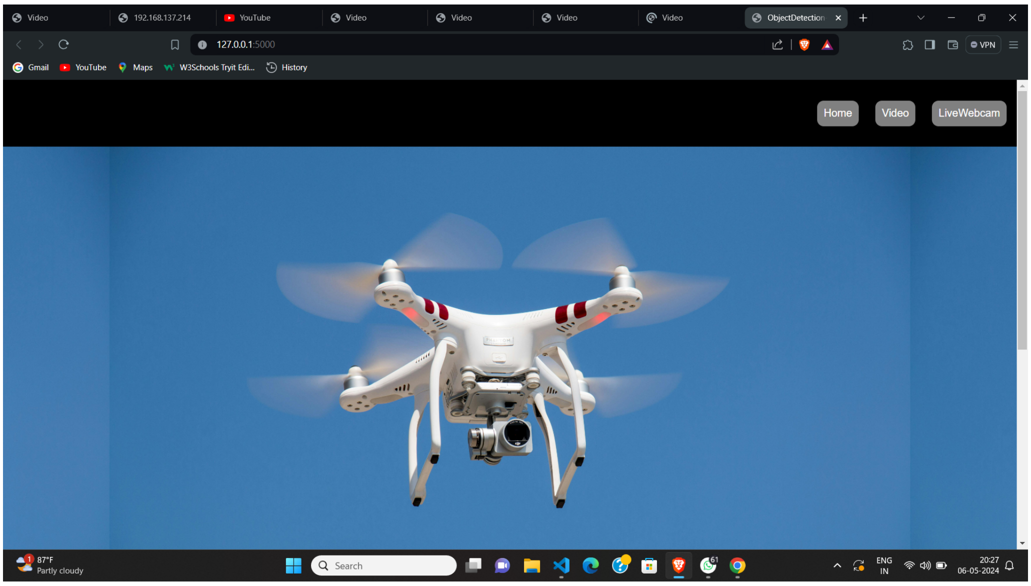


Fig. 10.1: Home page of the web application.

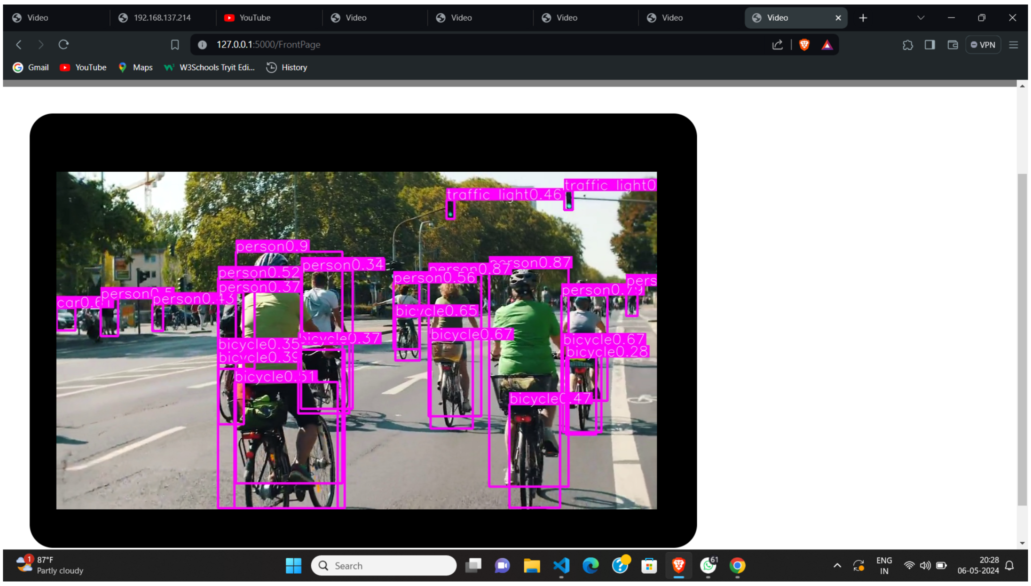


Fig. 10.2: Object detection from inputed video.

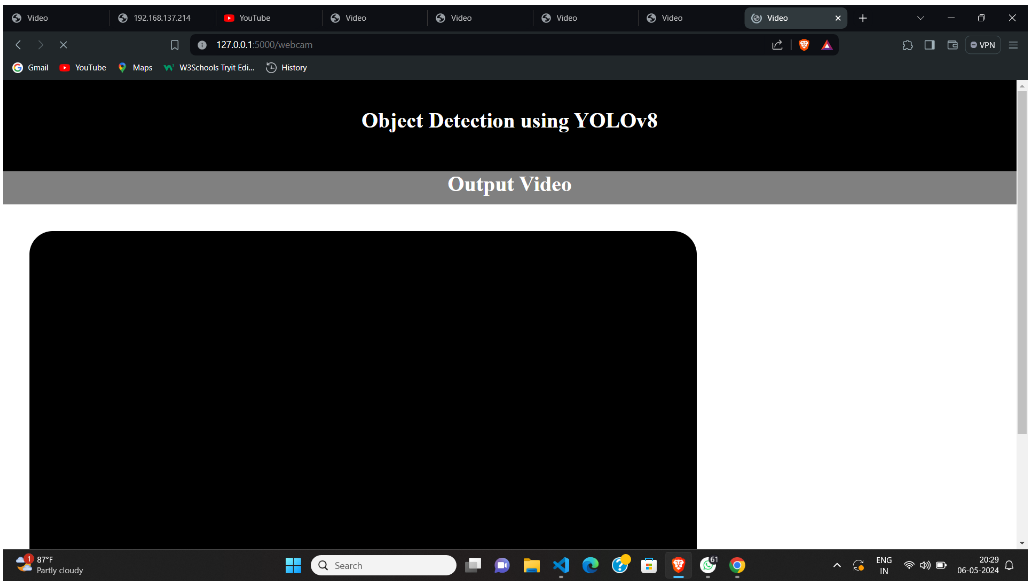


Fig. 10.3: Object detection interface for live drone footage.

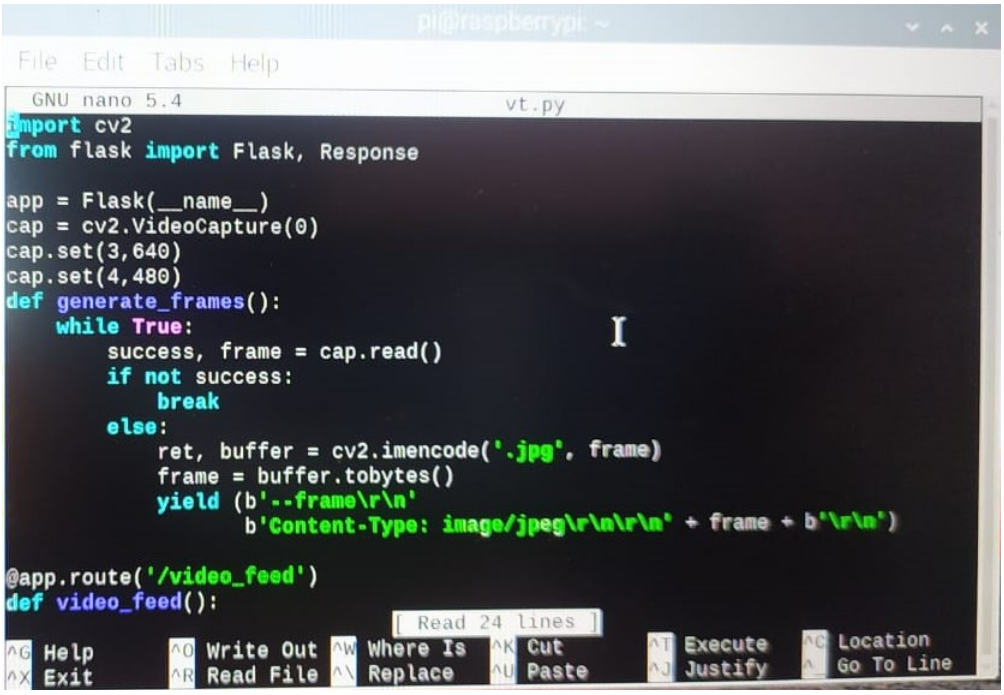
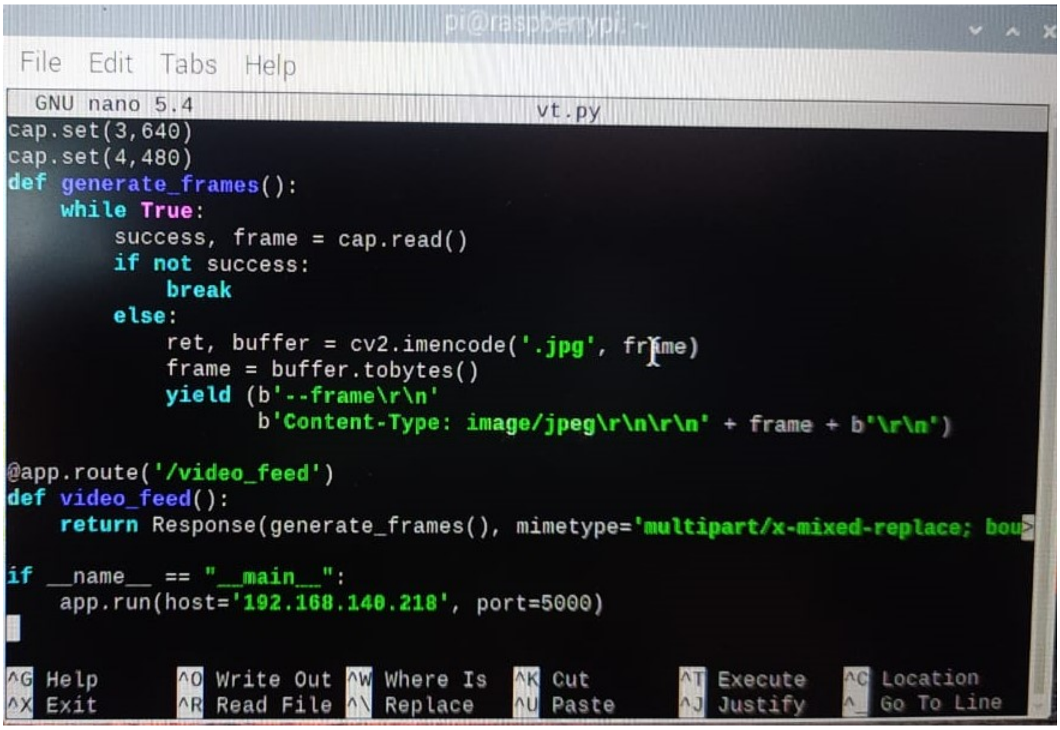


Fig. 10.4: Raspberry pi activation.



```
cap.set(3,640)
cap.set(4,480)
def generate_frames():
    while True:
        success, frame = cap.read()
        if not success:
            break
        else:
            ret, buffer = cv2.imencode('.jpg', frame)
            frame = buffer.tobytes()
            yield (b'--frame\r\n'
                   b'Content-Type: image/jpeg\r\n\r\n' + frame + b'\r\n')

@app.route('/video_feed')
def video_feed():
    return Response(generate_frames(), mimetype='multipart/x-mixed-replace; bou

if __name__ == "__main__":
    app.run(host='192.168.140.218', port=5000)
```

Fig. 10.5: Link generation to retrieve live footage.

APPENDIX-2

YOLO-V8 HUMAN DETECTION

from flask import Flask,
render$_{t}emplate, Response, jsonify, request, session$
from flask$_{w}tfimportFlaskForm$
from wtforms import
FileField,SubmitField,StringField,DecimalRangeField,IntegerRangeField
from werkzeug.utils import secure$_{f}ilename$
from wtforms.validators import InputRequired,NumberRange
import os
import cv2
from YOLO$_{V}ideoimportvideo_{d}etection$
app = Flask($_{name})$
app.config['SECRET$_{K}EY'] =' muhammadmoin'$
app.config['UPLOAD$_{F}OLDER'] =' static/files'$
#Use FlaskForm to get input video file from user
class UploadFileForm(FlaskForm):
#We store the uploaded video file path in the FileField in the
variable file
#We have added validators to make sure the user inputs the video in
the valid format and user does upload the
#video when prompted to do so
file = FileField("File",validators=[InputRequired()])
submit = SubmitField("Run")
def generate$_{f}rames(path_{x} ='') :$
yolo$_{o}utput = video_{d}etection(path_{x})$
for detection$_{i}nyolo_{o}utput$:
ref,buffer=cv2.imencode('.jpg',detection$_{)}$
frame=buffer.tobytes()
yield (b'–frame ◦ '
b'Content-Type: image/jpeg ◦ ◦ ' + frame +b' ◦ ')
def generate$_{f}rames_{w}eb(path_{x}) :$
yolo$_{o}utput = video_{d}etection(path_{x})$
for detection$_{i}nyolo_{o}utput$:
ref, buffer = cv2.imencode('.jpg', detection$_{)}$
frame = buffer.tobytes()
yield (b'–frame ◦ '
b'Content-Type: image/jpeg ◦ ◦ ' + frame + b' ◦ ')

```
@app.route('/', methods=['GET','POST'])
@app.route('/home', methods=['GET','POST'])
def home():
session.clear()
return render_template('indexproject.html')
# Rendering the Webcam Rage
#Now lets make a Webcam page for the application
#Use 'app.route()' method, to render the Webcam page at "/webcam"
@app.route("/webcam", methods=['GET','POST'])
def webcam():
session.clear()
return render_template('ui.html')
@app.route('/FrontPage', methods=['GET','POST'])
def front():
# Upload File Form: Create an instance for the Upload File Form
form = UploadFileForm()
if form.validate_on_submit():
# Our uploaded video file path is saved here
file = form.file.data
file.save(os.path.join(os.path.abspath(os.path.dirname(_file_)),
app.config['UPLOAD_FOLDER'],
secure_filename(file.filename))) Then
save the file
# Use session storage to save video file path
session['video_path'] =
os.path.join(os.path.abspath(os.path.dirname(_file_)),
app.config['UPLOAD_FOLDER'],
secure_filename(file.filename))
return render_template('videoprojectnew.html', form=form)
@app.route('/video')
def video():
#return Response(generate_frames(path_x='static/files/bikes.mp4'),
mimetype='multipart/x-mixed-replace; boundary=frame')
return Response(generate_frames(path_x = session.get('video_path',
None)),mimetype='multipart/x-mixed-replace; boundary=frame')
# To display the Output Video on Webcam page
@app.route('/webapp')
def webapp():
video_feed_url = "http://192.168.140.218:5000/video_feed"
```

#return Response(generate$_f rames(path_x = session.get('video_p ath',$
None),conf$_= round(float(session.get('conf_',$
None))/100,2)),mimetype='multipart/x-mixed-replace; boundary=frame')
return Response(generate$_f rames_w eb(path_x = video_f eed_u rl),$
mimetype='multipart/x-mixed-replace; boundary=frame')
if $_n ame_{==''}{}_m ain_{'':}$
app.run(debug=True)

VIDEO DETECTION

from ultralytics import YOLO
import cv2
import math
def video$_d etection(path_x)$:
video$_c apture = path_x$
#Create a Webcam Object
cap=cv2.VideoCapture(video$_c apture)$
frame$_w idth = int(cap.get(3))$
frame$_h eight = int(cap.get(4))$
#out=cv2.VideoWriter('output.avi', cv2.VideoWriter$_f ourcc('M','J',$
'P','G'), 10, (frame$_w idth, frame_h eight))$
model=YOLO("../YOLO-Weights/yolov8n.pt")
classNames = ["person", "bicycle", "car", "motorbike", "aeroplane",
"bus", "train", "truck", "boat",
"traffic light", "fire hydrant", "stop sign", "parking meter",
"bench", "bird", "cat",
"dog", "horse", "sheep", "cow", "elephant", "bear", "zebra",
"giraffe", "backpack", "umbrella",
"handbag", "tie", "suitcase", "frisbee", "skis",
"snowboard", "sports ball", "kite", "baseball bat",
"baseball glove", "skateboard", "surfboard", "tennis
racket", "bottle", "wine glass", "cup",
"fork", "knife", "spoon", "bowl", "banana", "apple",
"sandwich", "orange", "broccoli",
"carrot", "hot dog", "pizza", "donut", "cake", "chair",
"sofa", "pottedplant", "bed",
"diningtable", "toilet", "tvmonitor", "laptop", "mouse",
"remote", "keyboard", "cell phone",
"microwave", "oven", "toaster", "sink", "refrigerator",

”book”, ”clock”, ”vase”, ”scissors”,
”teddy bear”, ”hair drier”, ”toothbrush”]
while True:
success, img = cap.read()
results=model(img,stream=True)
for r in results:
boxes=r.boxes
for box in boxes:
x1,y1,x2,y2=box.xyxy[0]
x1,y1,x2,y2=int(x1), int(y1), int(x2), int(y2)
print(x1,y1,x2,y2)
cv2.rectangle(img, (x1,y1), (x2,y2), (255,0,255),3)
conf=math.ceil((box.conf[0]*100))/100
cls=int(box.cls[0])
class$_n ame = classNames[cls]$
label=f’class$_n ameconf'$
t$_s ize = cv2.getTextSize(label, 0, fontScale = 1,$
thickness=2)[0]
print(t$_s ize)$
c2 = x1 + t$_s ize[0], y1 - t_s ize[1] - 3$
cv2.rectangle(img, (x1,y1), c2, [255,0,255], -1,
cv2.LINE$_A A)\#filled$
cv2.putText(img, label, (x1,y1-2),0, 1,[255,255,255],
thickness=1,lineType=cv2.LINE$_A A)$
yield img
#out.write(img)
#cv2.imshow(”image”, img)
#if cv2.waitKey(1) 0xFF==ord(’1’):
#break
#out.release()
cv2.destroyAllWindows()

WEB DESIGNING

<!DOCTYPE html>
<html lang=”en”>
<head>
<meta charset=”UTF-8”>
<meta name=”viewport” content=”width=device-width, user-scalable=no,

```
initial-scale=1.0, maximum-scale=1.0, minimum-scale=1.0”>
<meta http-equiv=”X-UA-Compatible” content=”ie=edge”>
<title>ObjectDetection-YOLOv8/title>
body{ font-family:Helvetica, sans-serif; margin:0px; padding:0px; }
header.feature-box.right{ background-color:black; color:white; height:120px;
padding:5px; text-align:right; }
header.feature-box.below{ height:800px; background-color:blue;
margin-top:-30px; background-image:url(”../static/images/drone.jpg”);
background-size: contain; background-position:center; }
ul{ list-style:none; padding:0px;}
li{ display:inline-block; background-color:gray; border-radius:10px; padding:10px;
margin:10px; }
li:hover{ background-color:gray; margin:10px; }
header a{ color:white; text-decoration:none; width:100}
.details{ height:80px; margin:40px; padding:0px; font-size:40px;
background-color:gray; text-align:center; color:black; }
.features{ background:white; color:gray; display:flex; flex-direction:row;
padding:20px;
}
.features figure {/*background:pink;*/ width:400px; margin:auto; text-align:center;
text-transform:uppercase; }
.features figure img{ border-radius:10width:400px; }
div.main{ display:flex; margin:40px; letter-spacing:1px; h1, h2, h3, h4, h5, h6 {
font-family: ”Playfair Display”; letter-spacing: 5px; } }
.subpart2{ text-align:center; text-transform:uppercase; }
.subpart2 figure img{ border-radius:4width:800px; }
.contactpage margin:20px; }
textbox { font-size: 18pt; height: 42px; width : 1000px; }
textboxes { font-size: 18pt; height: 42px; width : 1000px; }
footer{ background-color:black; text-align:center; margin:0px; padding:0px;
color:gray; font-size:20px; }
</style>
</head>
<body>
<header class=”feature-box right”>
<nav>
<ul><li><a href=”/home”>Home</a></li><li><a
href=”/FrontPage”>Video</a></li><li><a
href=”/webcam”>LiveWebcam</a></li></ul>
```

```
</nav>
</header>
<header class=”feature-box below”>
</header>
<div class=”contactpage”>
<h1>Contact </h1><br>
<p >In case of any query do let me know </p>
<form action=”/action_page.php”target = ”_blank” >< p >< inputtype =
”text”placeholder = ”Name”requiredname = ”Name”id = ”textbox” >< /p ><
p >< inputtype = ”text”placeholder = ”EmailID”requiredname =
”EmailID”id = ”textboxes” >< /p >< p >< buttontype = ”submit” >
SENDMESSAGE < /button >< /p >
</form>
</div>
</body>
</html>
```

APPENDIX-3

ISSN: 2582-3930

ACCEPTANCE CERTIFICATE

Impact Factor: 8.448
DOI Prefix: 10.55041

INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING & MANAGEMENT (IJSREM)
An Open Access Scholarly Journal || Index in major Databases & Metadata

We are pleased to inform you that your Research Paper titled

AI Enabled Robot for Data Collection in Unreachable and Extreme Environments: A Review

has been ACCEPTED for publication in
INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING & MANAGEMENT (IJSREM)

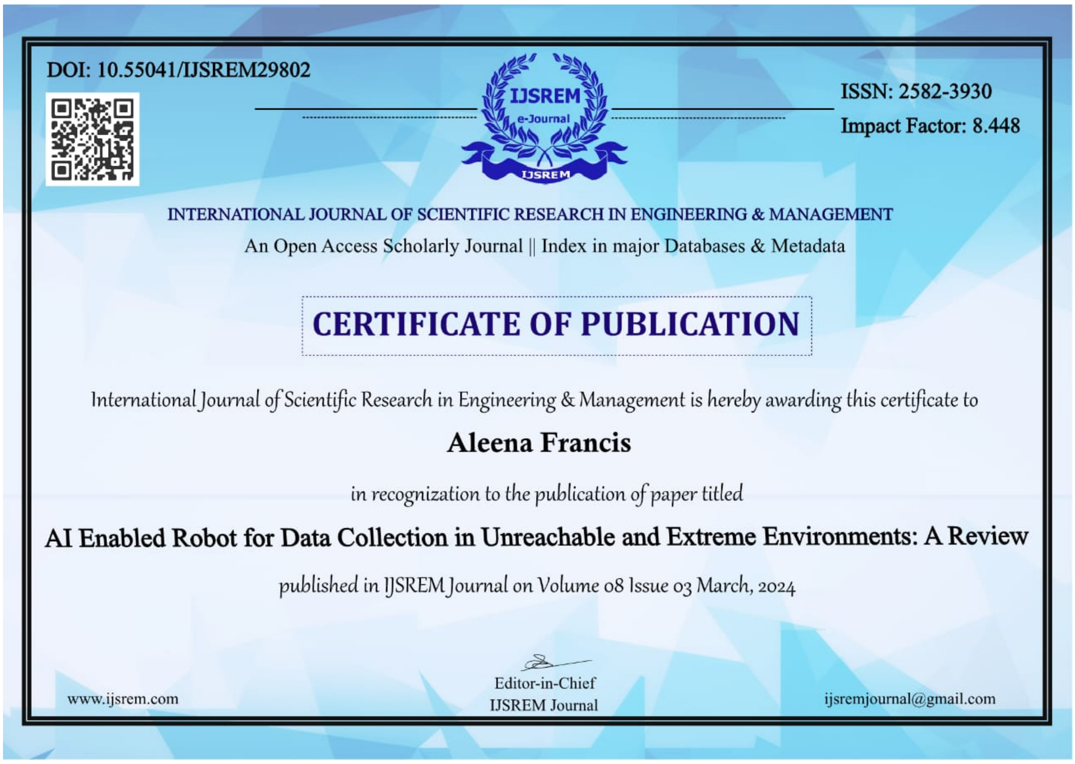
VOLUME 08 ISSUE 03 MARCH - 2024

We are delighted to see your commitment & hardwork to share your research is being recognized. We look foreward to helping you with all of your publication needs. Thank you for choosing IJSREM!!

Crossref CiteFactor

Editor-in-chief
IJSREM

www.ijsrem.com ijsremjournal@gmail.com



DOI: 10.55041/IJSREM29802

ISSN: 2582-3930
Impact Factor: 8.448

INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING & MANAGEMENT
An Open Access Scholarly Journal || Index in major Databases & Metadata

CERTIFICATE OF PUBLICATION

International Journal of Scientific Research in Engineering & Management is hereby awarding this certificate to

Aleena Francis

in recognization to the publication of paper titled

AI Enabled Robot for Data Collection in Unreachable and Extreme Environments: A Review

published in IJSREM Journal on Volume 08 Issue 03 March, 2024

Editor-in-Chief
IJSREM Journal

www.ijsrem.com ijsremjournal@gmail.com

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 03 | March - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

AI Enabled Robot for Data Collection in Unreachable and Extreme Environments: A Review

Aleena Francis[1], Amrutha Priya C B[2], Arjun K S[3], Divya P S[4], Dr. Nitha C Velayudhan[5]

[1,2,3,4]*Student, Department of computer science and engineering, Universal Engineering College, Kerala, India.*
[5]*HOD, Department of computer science and engineering, Universal Engineering College, Kerala, India.*

---***---

Abstract - This paper presents a groundbreaking approach to data collection in hazardous or inaccessible environments, presenting the design, development, and implementation of an innovative autonomous robot. The robot is designed to navigate and collect valuable data from locations too dangerous or remote for human exploration, enabling scientific research and exploration in unprecedented ways. The AI-powered drone is equipped for precise human identification, controlled through a user-friendly mobile app. The software analyzes live drone footage to detect human presence using models like YOLO, with high accuracy in real-time human detection tasks. The robot is equipped with an array of sensors, including cameras, and uses image processing technology for processing images. GPS tracking technology is used for device tracking. The proposed autonomous robot promises to revolutionize data collection in unreachable environments, opening new avenues for scientific discovery, resource assessment, and environmental monitoring.

***Key Words*:** YOLOV8, UAV, Python, Flask, Computer vision, AI.

1.INTRODUCTION

The "AI Enabled Robot for Data Collection in Unreachable and Extreme Environments" represents a groundbreaking fusion of artificial intelligence (AI) and robotics, marking a significant leap forward in technological innovation. Within the rapidly advancing field of drone technology, our system emerges as a trailblazing solution that seamlessly integrates sophisticated object detection capabilities using the YOLO (You Only Look Once) framework with intuitive mobile control functionalities.

At the core of our innovation lies the implementation of the YOLOv8 model, which showcases unparalleled proficiency in detecting human presence with exceptional precision and real-time efficiency. This model incorporates novel enhancements, such as a RepVGG structural reparametrized convolution module and an extended PAFPN structure, enabling the drone to acquire robust features without compromising on speed or model size. Consequently, our drone transcends the conventional role of an aerial vehicle, evolving into a sophisticated tool capable of delivering high-performance, multi-scale human presence detection.

Moreover, our drone system sets itself apart through its user-friendly mobile app interface, which adopts a customer-centric approach to drone control. Serving as a centralized command hub, the mobile app empowers users with effortless navigation capabilities, enabling them to seamlessly regulate the drone's movements. Additionally, the app provides real-time updates on detected human activity, fostering a dynamic feedback loop between the user and the drone. This seamless integration of YOLO's advanced object detection capabilities with the convenience of mobile control heralds the onset of a new era in Unmanned Aerial Vehicle (UAV) applications, with implications that extend beyond traditional surveillance and security.

The YOLOv8-powered drone boasts several key features that further enhance its efficacy. Firstly, its high accuracy stems from the utilization of advanced object detection algorithms, significantly minimizing the occurrence of false positives and negatives. Furthermore, the drone's adaptability to diverse environments, lighting conditions, and weather ensures consistent performance across various operational scenarios.

In essence, our innovation represents a paradigm shift in aerial presence detection and control, where cutting-edge technology converges to redefine the landscape of unmanned aerial systems. As we welcome this future characterized by unprecedented levels of efficiency, precision, and versatility, we anticipate transformative applications across a myriad of industries, catering to their evolving needs and challenges.

2. RELATED WORK

Ning Zhang, Francesco Nex, George Vosselman and Norman Kerle, [1] introduces human detection of images using deep learning has been a popular research topic in recent years and has achieved remarkable performance. Training a human detection network is useful for first responders to search for trapped victims in debris after a disaster. In this paper, we focus on the detection of such victims using deep learning, and we find that state-of-the-art detection models pretrained on the well-known COCO dataset fails to detect victims. This is because all the people in the training set are shown in photos of daily life or sports activities, while people in the debris. After a disaster, people usually only have parts of their bodies exposed. In addition, because of the dust, the colors of their clothes or body parts are similar to those of the surrounding debris. Compared with collecting images of common objects and images of disaster victims is extremely difficult training. Therefore, we propose a framework to generate harmonious composite images for training. We first paste body parts onto a debris background to generate composite victim images, and then use a deep harmonization network to make the composite images look more harmonious. We selectYOLOv5l as the most suitable model, and experiments show that using composite images for training improves the AP (average precision).

Ravindra R. Patil, Rajnish Kaur Calay, Mohamad Y. Mustafa, Saniya M. Ansari [2] introduces artificial intelligence (AI) uses computer vision models to interpret and recognize the visual world, similar to human vision. This technology relies on extensive data and human expertise to yield accurate results. However, locating and resolving blockages in sewer systems is a complex task due to their diverse nature and lack of robust techniques. This research uses the "S-BIRD" dataset as the foundation for a deep neural network model, with transfer learning and fine-tuning techniques applied on the YOLOv5 architecture. The trained model achieves a remarkable accuracy rate in sewer blockage detection, enhancing the reliability and efficacy of the robotic framework for efficient blockage removal. The model achieved a mean average precision score of 96.30% at a confidence threshold of 0.5, maintaining a consistently high-performance level of 79.20% across Intersection over Union (IoU) thresholds.

Peng Zhang, Weimin Lei, Xinlei Zhao, Lijia Dong and Zhaonan Lin, [3] presents crowd counting is a crucial task in fields like video surveillance, accident prediction, public security, and intelligent transportation. However, it faces challenges such as large-scale crowd aggregation in public places, positioning errors in large-scale datasets, and inconsistent human head target size in dense images. Existing crowd counting methods mainly use density plot regression methods, which do not distinguish between distant and near targets and cannot adaptively respond to scale changes. To address these issues, an adaptive multi-scale far and near distance network based on the convoluted neural network (CNN) framework is proposed. The model uses stacked convolution layers to deepen the network's depth, allocate different receptive fields based on the distance between the target and the camera, and fuse features between nearby targets to enhance pedestrian feature extraction. Depth information is used to distinguish distant and near targets of different scales, and the original image is cut into four different patches for pixel-level adaptive modelling. Density normalized average precision (nAP) indicators are added to analyses the method's accuracy in spatial positioning.

Zhengxin Zhang, [4] the article proposes Drone-YOLO, a series of multi-scale UAV image object detection algorithms based on the YOLOv8 model, to overcome challenges in UAV imagery. The algorithms include a three-layer PAFPN structure, a detection head for small-sized objects, and a sandwich-fusion module. They also use RepVGG modules as down sampling layers. The Drone-YOLO methods have been evaluated on the VisDrone2019 dataset and show significant improvements in object detection accuracy. The parameter-efficient Drone-YOLO (tiny) performs equivalently or better than the baseline method with 9.66M parameters, proving the effectiveness of the methods in drone image object detection.

M.D. Mursalin and Syed Mohammed Shamsul Islam, [5] they introduce this study proposes a pipeline for automated ear detection from 3D profile face images, focusing on semantic part segmentation. The ear detection problem is formulated as a semantic part segmentation problem, detecting the ear directly in 3D point clouds of profile face data. The proposed pipeline includes synthetic data generation and ground-truth data labelling. EarNet, a modified version of the PointNet++ architecture, is introduced to handle pose variations in real data. An automatic tool is developed to create groundtruth labels of any 3D public data set, including co-registered 2D images. The experimental result show higher localization compared to existing methods.

3. DESIGN AND ANALYSIS

3.1. TRAINING A DISASTER VICTIM DETECTION NETWORK FOR UAV SEARCH AND RESCUE USING HARMONIOUS COMPOSITE IMAGES

The research focuses on improving human detection in disaster scenarios using deep learning techniques. Existing models trained on standard datasets like COCO fail to detect disaster victims due to differences in context and appearance. Victims in debris are often only partially visible, and their colors blend with the surrounding debris, making them challenging to detect. To address this, the paper proposes a framework to generate composite images by combining human body parts with debris backgrounds. These synthetic images are crucial for training due to the difficulty in obtaining real victim images, undergo a process using a deep harmonization network.to enhance their realism. We select YOLOv5l as the most suitable model, and moments show that using composite images for training. This approach significantly boosts detection accuracy, demonstrated by a 19.4% improvement in average precision (AP). Additionally, employing these harmonious images further enhances the victim detection model, resulting in a total AP increase of 10.2%. The research con-tributes by introducing a method to create synthetic training data specifically tailored for detecting human body parts in disaster debris, critical for post-disaster search and rescue operations. The integration of a deep harmonization network significantly enhances the credibility and effectiveness of these synthetic images in training the detection model, offering promise for real-time deployment on unmanned aerial vehicles (UAVs) for autonomous post-disaster search missions. The illustration of proposed system is shown in Fig-1.

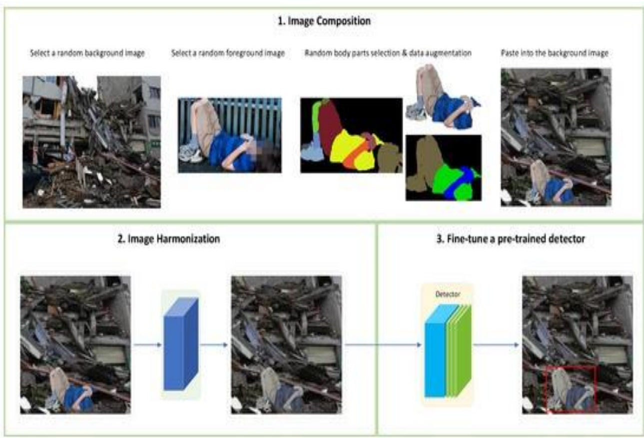


Fig -1: Proposed framework.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 03 | March - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

3.2. AI-DRIVEN HIGH-PRECISION MODEL FOR BLOCKAGE DETECTION IN URBAN WASTE WATER SYSTEMS

The research focuses on field of artificial intelligence (AI), computer vision plays a crucial role by simulating human vision to intelligently interpret and recognize visual information. Similar to human perception, this technology depends on integrating extensive datasets and human expertise in a meticulously structured manner, resulting in precise and accurate outcomes. The challenging task of identifying and addressing obstructions within sewer systems, due to their diverse nature and the lack of robust techniques, has spurred innovative research. This research utilizes the "S-BIRD" dataset, a comprehensive collection of frames depicting sewer blockages, as the foundational training data for a deep neural network model. The YOLOv5 architecture serves as the framework, and the strategic implementation of transfer learning and fine-tuning techniques using the corresponding dataset enhances the model's performance, leading to optimal outcomes. The trained model exhibits exceptional accuracy in detecting sewer blockages, thereby enhancing the reliability and efficiency of the associated robotic framework designed for the proficient removal of diverse blockages. Noteworthy is the achieved mean average precision (mAP) score of 96.30% at a confidence threshold of 0.5, maintaining a consistently high-performance level of 79.20% across Intersection over Union (IoU) thresholds ranging from 0.5 to 0.95. This research represents a significant advancement in AI-driven solutions for modern urban sanitation systems. Anticipated contributions include a substantial improvement in the effectiveness of sewer maintenance and a broader application of AI technologies to address challenges within urban infrastructure. The illustration of architectural perception of YOLOv5 model is shown in Fig-2.

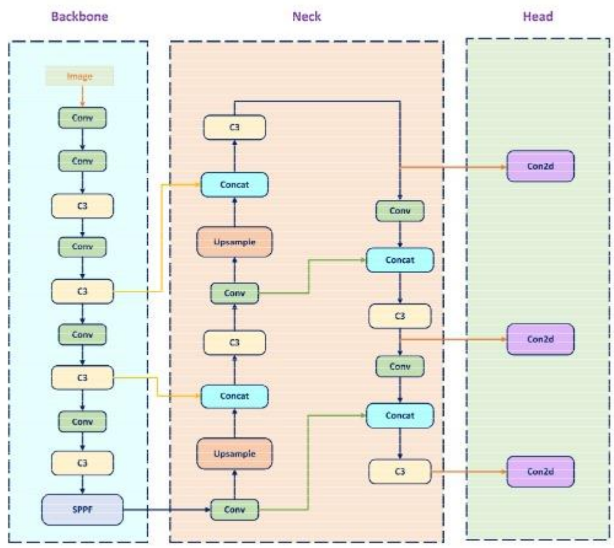


Fig -2: Architectural perception of YOLOv5 model.

3.3. AN ADAPTIVE MULTI-SCALE NETWORK BASED ON DEPTH INFORMATION FOR CROWD COUNTING

This presents the task of crowd counting holds crucial significance in several domains, including video surveillance, accident prediction, public security, and intelligent transportation. However, it encounters several challenges. Firstly, high-density crowd aggregation in public places often leads to severe occlusions, complicating accurate counting. Errors in annotating large-scale datasets also impact training outcomes. Moreover, in densely populated images, the sizes of human head targets vary widely, making it challenging for a single network to identify both near and far targets concurrently. Current crowd counting methods, typically relying on density plot regression, struggle to differentiate features between distant and near targets, limiting their adaptability to scale changes and impacting detection performance in sparser population areas. To address these issues, a solution is proposed a novel adaptive multi-scale network based on convolutional neural networks (CNN). This network aims to count dense populations while achieving a balance between accuracy, inference speed, and overall performance. To enable the model to distinguish between near and far features, stacked convolution layers are employed to deepen the network's depth. Different receptive fields are allocated based on the distance between targets and the camera, enhancing feature extraction for nearby pedestrians. Depth information is utilized to differentiate between distant and near targets of various scales. Additionally, the original image is divided into four patches for pixel-level adaptive modelling of the population. Furthermore, density normalized average precision (nAP) indicators are introduced to assess spatial positioning accuracy. The proposed NF-Net is validated on challenging benchmarks, including Shanghai Tech Part A and B, UCF CC 50, and UCF-QNRF datasets. Comparative analysis against state-of-the-art methods demonstrates its superior performance across diverse scenarios. Notably, on the UCF-QNRF dataset, the method effectively resolves complex background interference, showcasing its efficacy in countering intricate challenges in crowd counting tasks. The illustration of proposed system is shown in Fig-3.

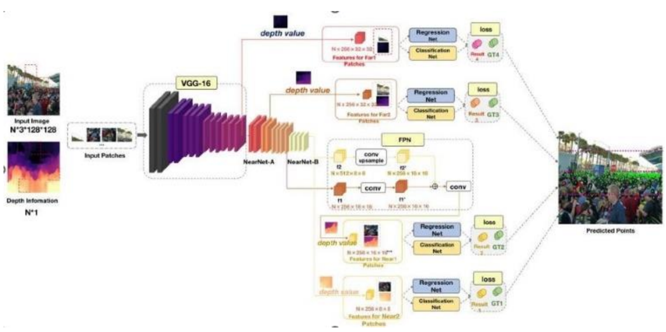


Fig -3: The overall framework of the proposed NF-Net. The backbone of NF-Net adopts all convolutional layers in VGG 16 and a NearNet network.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 03 | March - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

3.4. DRONE-YOLO: AN EFFICIENT NEURAL NETWORK METHOD FOR TARGET DETECTION IN DRONE IMAGES

The article introduces Drone-YOLO, a set of advanced object detection algorithms specially tailored for detecting objects in unmanned aerial vehicle (UAV) imagery, which presents distinct challenges such as large image sizes, small-sized objects, dense distribution, overlapping instances, and inadequate lighting. To overcome these hurdles, the Drone-YOLO algorithms, based on the YOLOv8 model, propose several enhancements, particularly focusing on the neck component of the model. They incorporate a three-layer PAFPN structure and a specialized detection head for small-sized objects using large-scale feature maps, significantly improving the algorithm's capability to detect smaller targets in UAV imagery. Moreover, they introduce the sandwich-fusion module into each layer of the neck's up-down branch, combining network and low-level features to provide detailed spatial information across different layers. This fusion utilizes depth wise separable evolution, balancing parameter costs and a broader receptive field. Additionally, the algorithms utilize RepVGG modules in the network backbone, enhancing the learning of multi-scale features and surpassing traditional convolutional layers. Evaluation on the VisDrone2019 dataset demonstrates that the proposed Drone YOLO (large) outperforms other baseline methods in object detection accuracy, exhibiting substantial improvements in mAP0.5 metrics compared to YOLOv8. Notably, the parameter-efficient Drone-YOLO (tiny) with fewer parameters performs comparably or even better than methods with higher parameter counts on the dataset, validating the efficacy of Drone-YOLO algorithms for object detection in drone imagery. The illustration of proposed network model is shown in Fig-4.

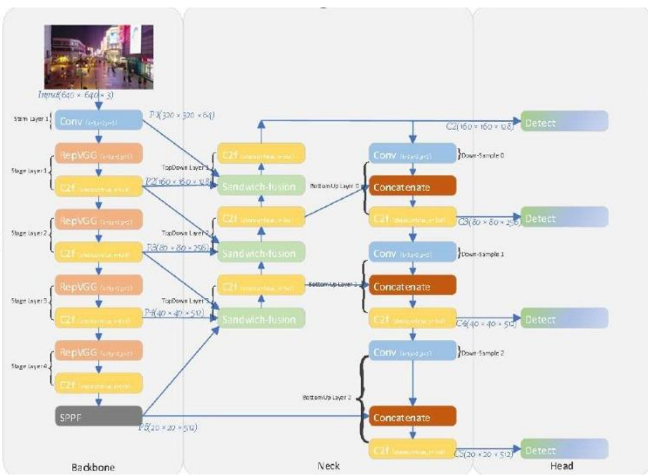


Fig- 4: Proposed network model (large) with its backbone, neck and head.

3.5. DEEP LEARNING FOR 3D EAR DETECTION

They introduce delves into utilizing the unique features of human ear shapes for various applications, such as biometric identification, 3D ear reconstruction, gender recognition, genetic studies, and clinical asymmetry analysis. While most ear detection methods rely on 2D images, the paper highlights the limitations of these approaches in handling unconstrained scenarios due to sensitivity to lighting and pose variations. To overcome these limitations, the study proposes leveraging 3D images and explores the advancements in 3D imaging techniques that have accelerated applications in biometrics, robotics, medical diagnosis, and autonomous driving. However, conventional convolutional neural networks (CNNs) face challenges in directly processing regularly ordered 3D point clouds, prompting the introduction of novel architectures like PointNet and PointNet++ that can analyse features directly on 3D point clouds. These advancements have paved the way for solving various research queries in classification and semantic segmentation tasks within the 3D domain. The research introduces EarNet, a modified version of PointNet++, specifically designed to detect ears directly from 3D point clouds of profile face data. To address pose variations, a rotation augmentation block is included during transfer learning of EarNet. Since labelled 3D point cloud data for ear detection is scarce, the study proposes a novel method for generating synthetic 3D profile face data to train EarNet. Furthermore, a method is devised to create ground-truth labels on real 3D data for quantitative evaluation of EarNet's performance. The contributions of the work encompass the proposal of EarNet, a modified deep learning model, the synthetic generation of 3D profile face data for training, a method for creating ground-truth labels on real 3D data, and comprehensive experiments showcasing state-of-the-art performance on the largest publicly available 3D profile face dataset. The illustration of block diagram of ground-truth labeling procedure on real data is shown in Fig-5.

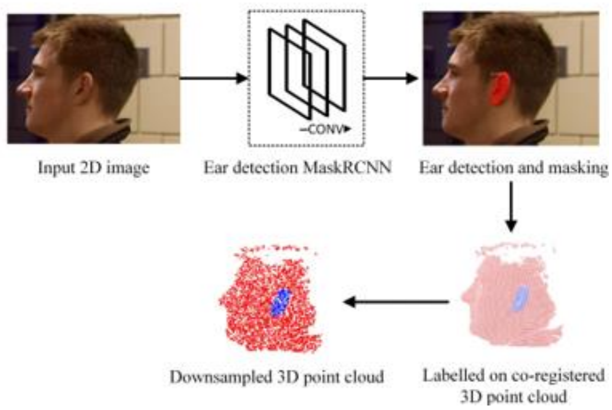


Fig -5: Block diagram of ground-truth labeling procedure on real data.

4. DISCUSSION

The drone system designed for human presence detection is a highly sophisticated and advanced technological solution that leverages the cutting edge YOLOv8 object detection algorithm and is seamlessly controlled through a dedicated mobile application. The integration of the YOLOv8 model empowers the drone to efficiently and accurately detect human presence in real-time, a capability that is crucial for applications such as security, surveillance, and search and rescue missions. A key component of the system is its onboard camera, which captures live video feed that is then processed in real-time by the YOLOv8 algorithm for instantaneous object detection, with a specific focus on identifying humans within the captured imagery. This seamless integration with a user-friendly mobile application not only enables users to control the drone's operations remotely but also provides them with the ability to monitor detection results effectively.

This combination of features makes the drone system highly versatile and adaptable for a wide range of applications, enhancing its usability and effectiveness in various operational scenarios. Moreover, the system incorporates an advanced object detection algorithm based on YOLOv8, incorporating significant enhancements to further improve its performance and accuracy. The neural network architecture underlying the system is meticulously designed for human presence detection, especially in the context of Unmanned Aerial Vehicle (UAV) applications. This architecture includes several key components and concepts, such as building upon the YOLOv8-l model, integrating the RepVGG structural re-parameterized convolution module, and implementing the PAFPN structure (Path Aggregation Feature Pyramid Network) in the neck section of the neural network.

These components are strategically designed to enhance the network's capability in detecting relatively small objects, particularly humans, from UAV imagery. The network is positioned as a sophisticated and advanced solution explicitly tailored for UAV applications, emphasizing its ability to ensure real-time, accurate, and multi-scale human presence detection from UAV imagery. This strategic positioning highlights the system's adaptability, efficiency, and effectiveness in addressing complex operational challenges and underscores its significance as a cutting-edge technological solution in the field of UAV applications

5. CONCLUSIONS

An innovative autonomous robot has been designed and implemented to collect essential data in hazardous or inaccessible environments. The robot is engineered to navigate and collect data from locations too dangerous or remote for human exploration, unlocking unprecedented possibilities for scientific research and exploration. The AI powered drone, capable of precise human identification through a user-friendly mobile control interface, leverages advanced deep learning approaches and architectures like YOLO for remarkable accuracy in real-time human detection tasks. The robot is equipped with a diverse array of sensors, including cameras, and uses advanced image processing and GPS tracking technologies for thorough data collection and processing. The integration of a notification system for critical situations enhances the robot's responsiveness in challenging environments. This technological amalgamation holds the promise of revolutionizing data collection in previously considered unreachable environments, opening new avenues for scientific discovery, resource assessment, and environmental monitoring. The proposed autonomous robot is a catalyst for transformative advances in scientific exploration, resource evaluation, and environmental surveillance in remote and dangerous locations.

SOME OF ADVANTAGES

a) Accessibility to hazardous or remote areas where human presence is limited or unsafe.
b) Adaptability to navigate complex terrains and adjust to changing environmental conditions.
c) Minimization of risks to human life by autonomously handling dangerous tasks in extreme environments.
d) AI analysis of collected data provides valuable insights and trends for informed decision-making and enhanced understanding of the environment.
e) Optimization of data collection routes for efficient and effective data gathering.

ACKNOWLEDGEMENT

The authors extend sincere gratitude to Dr.Nitha C Velayudhan for her Guidance and project coordinators, Mrs. Najla Nazar and Mrs. Gishma K M for mentorship. Their expertise played a vital role in shaping the direction and focus of this research.

REFERENCES

[1]. Ning Zhang, Francesco Nex, George Vosselman, Norman Kerle, “Training a disaster victim detection network for UAC search and rescue using harmonious composite images”. 2022.
[2]. Ravindra. R, Patil, Rajnish Kaur Calay, Mohammad. Y, Mustafa, Saniya. M, Ansari, “AI-Driven High-Precision Model for Blockage Detection in Urban Wastewater Systems”. 2023.
[3]. Zhengxin Zhang, “Article Drone-YOLO: An Efficient Neural Network Method for Target Detection in Drone Images”. 2023.
[4]. Peng Zhang, Weimin Lei, Xinlei Zhao, Lijia Dong and Zhaonan Lin, “An Adaptive Multi-Scale Network Based on Depth Information for Crowd Counting”. 2023.

[5]. M.D. Mursalin and Syed Mohammed Shamsul Islam, “Deep Learning for 3D Ear Detection: A Complete Pipeline from Data Generation to Segmentation”.
2021.

Certificate

This is to certify that

Aleena Francis

presented a paper titled

AI Enabled Robot for Data Collection in Unreachable and Extreme Environments

in

NACORE NATIONAL CONFERENCE ON RESEARCH IN EMERGING AREAS 2024

organized by Department of Computer Science & Engineering, Amal Jyothi College of Engineering, Kanjirappally, co-sponsored by KSCSTE (Kerala State Council for Science, Technology & Environment) in association with ACM Kottayam Professional Chapter from 22-24 April 2024.

Dr Lillykutty Jacob
Principal, Amal Jyothi

Dr Juby Mathew
HOD, CSE, General Chair

Prof Manoj T Joy
Chair, ACM Kottayam Chapter

Prof Syam Gopi
Organizing chair, NACORE 24

KSCSTE

acm Kottayam Chapter

AMAL JYOTHI COLLEGE OF ENGINEERING KANJIRAPPALLY, KOTTAYAM, KERALA

AMAL JYOTHI
COLLEGE OF ENGINEERING
AUTONOMOUS
KANJIRAPPALLY

International Journal on Emerging Research Areas (IJERA) *ISSN2230-9993*

DOI:

AI Enabled Robot for Data Collection in Unreachable and Extreme Environments

1st Amrutha Priya C B
Student, Dept of Computer Science and Engineering
Universal Engineering College
Thrissur,Kerala
amruthapriyacb@gmail.com

2nd Dr. Nitha C Velayudhan
Professor, Dept of Computer Science and Engineering
Universal Engineering College
Thrissur, Kerala
nithacvelayudhan@uec.ac.in

3rd Arjun K S
Student, Dept of Computer Science and Engineering
Universal Engineering College
Thrissur,Kerala
arju4arju4@gmail.com

4th Aleena Francis
Student, Dept of Computer Science and Engineering
Universal Engineering College
Thrissur,Kerala
aleenafrancis2510@gmail.com

5th Divya P S
Student, Dept of Computer Science and Engineering
Universal Engineering College
Thrissur,Kerala
divyaachu2002@gmail.com

***Abstract*—This article outlines a groundbreaking approach to gathering data in hazardous or inaccessible environments through the utilization of innovative robotics. These robots are specifically designed to navigate and collect vital information from areas too dangerous or remote for human exploration, enabling unprecedented research opportunities. Central to this advancement is the integration of artificial intelligence (AI) support within drones, endowed with human recognition capabilities . By analyzing live drone footage using advanced pattern recognition techniques like YOLO (You Only Look Once), these drones achieve high-precision, real-time human detection. Equipped with an array of sensors, including cameras and GPS tracking systems, these autonomous robots are poised to revolutionize data collection and analysis in challenging environments. The proposed drone system represents a stateof-the-art solution to object detection challenges in harsh settings. By amalgamating cutting-edge technologies such as GPS tracking, obstacle avoidance, altitude holding features, and the YOLOv8 algorithm, this system offers unparalleled real-time monitoring and situational awareness capabilities. Leveraging GPS monitoring for efficient object localization and the YOLOv8 algorithm for quick and accurate detection, coupled with the drone's adeptness at navigating difficult terrain and maintaining stable flight, ensures consistent and dependable video feed quality. Moreover, a comprehensive strategy is employed to enhance safety by mitigating potential hazards while simultaneously boosting operational efficiency. This drone system holds promise for the delivering of the exceptional performance and invaluable insights in the face of challenging circumstances, whether deployed for environmental monitoring, surveillance missions, or search and rescue operations. The methodology for object detection using YOLOv8 involves a series of steps including pre-processing the input video, running the object detection model, initializing object post-processing, detecting objects over the frame, periodically re-detecting objects, and visualizing the results. Testing was conducted using the COCO dataset, which encompasses various lighting conditions, with datasets divided into testing, validation, and training categories to ensure robust performance evaluation. Photos with a resolution of 640 × 640 were utilized for experimentation, underscoring the efficacy of the proposed approach in addressing object detection challenges across diverse environmental conditions.**

***Index Terms*—YOLOv8, UAV, python, Flask, Computer vision, AI.**

I. Introduction

Object detection in images and videos refers to the process of using computer vision to identify and locate specific objects in images or videos. This function is important in many applications such as autonomous driving, surveillance, medical and virtual reality. YOLO (You Only Look Once) model has gained popularity in recent years thanks to its integration and accessories. End-to-end optimization makes them a solution for search operations. Unlike traditional object detection methods that handle multiple steps such as localization, elimination, and classification, YOLO's model performs these operations simultaneously in a single neural network, making the search faster and better. Main advantages of YOLO model is characterized by the ability to provide high results during target detection. YOLO provides accurate localization and classification of objects in the image by dividing the input image into grids and predicting the bounding box and classes that will appear in each grid cell.

AI robots collecting data in ultra-poor environments represents a groundbreaking combination of artificial

DOI:

intelligence (AI) and robots and marks a major leap forward in technology innovation. In the rapid development of drone technology, our system has evolved into a solution that seamlessly integrates complex object detection capabilities using the YOLO (Pick One) framework with intuitive motion control capabilities. This includes using the YOLOv8 standard, which has In addition, the application provides instant updates on the detection of human activity, creating positive feedback between the user and the drone.

YOLO's perfect combination of object detection capabilities and easy control opens a new era in unmanned aerial vehicle (UAV) applications with the high impact of normal surveillance and security.

YOLOv8 powered drones have several important features that strengthen their good work. First of all, its high accuracy is due to the use of the highest target to detect algorithms that reduce the likelihood of false alarms and negative effects. Additionally, the drone's adaptability to different environments, lighting, and weather conditions ensures consistent performance in a variety of operational situations. Fundamentally, our innovation represents a paradigm shift in detection and control where technologies can be combined to transform the UAS environment. As we embrace the future of quality, efficiency and diversity, we expect practice changes to occur across multiple industries to meet these changes and challenges.

II. Related Works

Ning Zhang, Francesco Nex, George Vosselman and Norman Kerle, [1] introduces human detection of images using deep learning has been a popular research topic in recent years and has achieved remarkable performance. Training a human detection network is useful for first responders to search for trapped victims in debris after a disaster. In this paper, we focus on the detection of such victims using deep learning, and we find that state-of-the-art detection models pretrained on the well-known COCO dataset fails to detect victims. This is because all the people in the training set are shown in photos of daily life or sports activities, while people in the debris. After a disaster, people usually only have parts of their bodies exposed. In addition, because of the dust, the colors of their clothes or body parts are similar to those of the surrounding debris. Compared with collecting images of common objects and images of disaster victims is extremely difficult training. Therefore, we propose a framework to generate harmonious composite images for training. We first paste body parts onto a debris background to generate composite victim images, and then use a deep harmonization network to make the composite images look more harmonious. We selectYOLOv5l as the most suitable model, and experiments show that using composite images for training improves the AP (average precision).

Ravindra R. Patil, Rajnish Kaur Calay, Mohamad Y. Mustafa, Saniya M. Ansari [2] introduces artificial intelligence (AI) uses computer vision models to interpret and recognize the visual world, similar to human vision. This technology relies on extensive data and human expertise to yield accurate results. However, locating and resolving blockages in sewer systems is a complex task due to their diverse nature and lack of robust techniques. This research uses the "S-BIRD" dataset as the foundation for a deep neural network model, with transfer learning and fine-tuning techniques applied on the YOLOv5 architecture. The trained model achieves a remarkable accuracy rate in sewer blockage detection, enhancing the reliability and efficacy of the robotic framework for efficient blockage removal. The model achieved a mean average precision score of 96.30% at a confidence threshold of 0.5, maintaining a consistently high-performance level of 79.20% across Intersection over Union (IoU) thresholds.

Peng Zhang, Weimin Lei, Xinlei Zhao, Lijia Dong and Zhaonan Lin, [3] presents crowd counting is a crucial task in fields like video surveillance, accident prediction, public security, and intelligent transportation. However, it faces challenges such as large-scale crowd aggregation in public places, positioning errors in large-scale datasets, and inconsistent human head target size in dense images. Existing crowd counting methods mainly use density plot regression methods, which do not distinguish between distant and near targets and cannot adaptively respond to scale changes. To address these issues, an adaptive multi-scale far and near distance network based on the convoluted neural network (CNN) framework is proposed. The model uses stacked convolution layers to deepen the network's depth, allocate different receptive fields based on the distance between the target and the camera, and fuse features between nearby targets to enhance pedestrian feature extraction. Depth information is used to distinguish distant and near targets of different scales, and the original image is cut into four different patches for pixel-level adaptive modelling. Density normalized average precision (nAP) indicators are added to analyses the method's accuracy in spatial positioning.

Zhengxin Zhang, [4] the article proposes Drone-YOLO, a series of multi-scale UAV image object detection algorithms based on the YOLOv8 model, to overcome challenges in UAV imagery. The algorithms include a three-layer PAFPN structure, a detection head for small-sized objects, and a sandwich-fusion module. They also use RepVGG modules as down sampling layers. The Drone-YOLO methods have been evaluated on the VisDrone2019 dataset and show significant improvements in object detection accuracy. The parameter efficient Drone-YOLO (tiny) performs equivalently or better than the baseline method with 9.66M parameters, proving the effectiveness of the methods in drone image object detection. M.D. Mursalin and Syed Mohammed Shamsul Islam, [5] they introduce this study proposes a pipeline for automated ear detection from 3D profile face images, focusing on semantic part segmentation. The ear detection problem is formulated as a semantic part segmentation problem, detecting the ear directly in 3D point clouds of profile face data. The proposed pipeline includes synthetic data generation and ground-truth data labelling. EarNet, a modified version of the PointNet++ architecture, is introduced to handle pose variations in real data. An automatic

DOI:

tool is developed to create ground truth labels of any 3D public data set, including co-registered 2D images. The experimental result show higher localization compared to existing methods.

III. Proposed System

The proposed drone system is a state-of-the-art remedy made to address object detecting problems in harsh settings. With the integration of cutting-edge technologies including GPS tracking, obstacle avoidance, altitude holding features, and the YOLOv8 algorithm, this system provides unmatched real-time monitoring and situational awareness capabilities. While GPS monitoring offers reliable location data to enable efficient object localization, the YOLOv8 algorithm guarantees quick and accurate detection of things in the drone's vicinity. The drone's ability to handle difficult terrain and maintain stable flight, along with the presence of obstacle avoidance sensors and altitude holding mechanisms, ensure consistent and dependable video feed quality. This all-encompassing strategy improves safety by reducing potential hazards while simultaneously increasing operational efficiency. This drone system promises to provide outstanding performance and insightful data in the most trying circumstances, whether it is used for environmental monitoring, surveillance missions, or search and rescue activities.

Pre-processing the input video, running the object detection model, initializing object postprocessing, detecting the objects over the frame, periodically re-detecting objects,and visualizing the results are all part of the suggested method for object detection using YOLOv8. The COCO dataset, which includes both natural daylight and various low-light photos, was employed in this experiment. Following preparation, the datasets were split into three categories: testing, validation, and training. We tested our suggested algorithms on photos with 640 × 640 resolution. Figure 1 illustrates the video detection procedure.

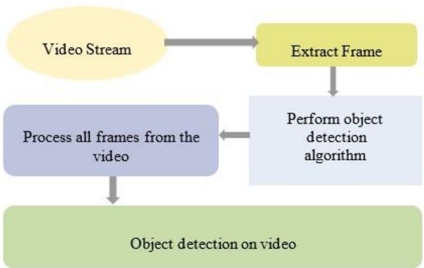


Fig. 1. Process of object detection on video.

We gave each algorithm 100 epochs and 16 batch sizes throughout the training phase. Following training, we obtained the following results for YOLOv8: f1-score 0.86%, precision 93.9%, recall 98%, and mAP 91.2%. Optimize the model's performance by adjusting its hyperparameters or by training it on a dataset that is more evenly distributed. Following the completion of the training phase, we conducted testing. The validation sets had 50 images, while the testing image set included 40 images.

A. *Preprocessing*

The proposed technique is an object detection model that recognizes objects in images and videos using deep learning algorithms. Preparing the unprocessed input image for feeding into the neural network is the pre-processing stage in models.

B. *Features extraction*

Our proposed models work with features by using a deep CNN to extract properties from the raw image. The CNN is made up of numerous convolutional layers that filter the original image in multiple ways before activation functions introduce non-linearity to the model. Every convolutional layer captures a different level of abstraction, ranging from simple features like corners and edges to more intricate aspects like shapes and textures. The last convolutional layer creates the high-level feature maps, which encode information about the image's content. Once features at different scales have been collected using the feature pyramid architecture, models employ a set of anchor boxes to forecast the exact location and size of objects in the image.

C. *Object detection on video*

To start object detection in video using the YOLOv8 model, the video frames are fed into the model. The model then generates object proposals or regions of interest (ROIs) in the input frame using a sliding window technique. These ROIs are selected according to how likely they are to contain an object. The model then uses convolutional neural networks (CNNs) to extract information from the ROIs. These attributes are used to identify the items that are present in the ROIs. The found items are then classified using softmax regression, which produces a probability distribution over the chosen object categories. Non-maximum suppression (NMS) is used by the model on the object proposals in order to improve detection accuracy and get rid of duplicate detections. By doing this, the boundary boxes are eliminated. This step removes bounding boxes with lower confidence scores when multiple bounding boxes overlap the same item. Bounding boxes and confidence ratings for every object found in the video are also included in the output frame.

IV. Results and Discussion

In this section, the detection results are shown in Figure 2. Our proposed algorithms detected the object in videos and achieved higher detection accuracy. In the below figure (a, b, c,d,e,f) the detected images achieved from the YOLOv8.

For the purpose of training, the deep learning algorithm require labelled data in the.txt format. Following annotation, we produced a.yaml file with the location of the image and the number of classes. One hundred epochs have carried out the suggested trials.

DOI:

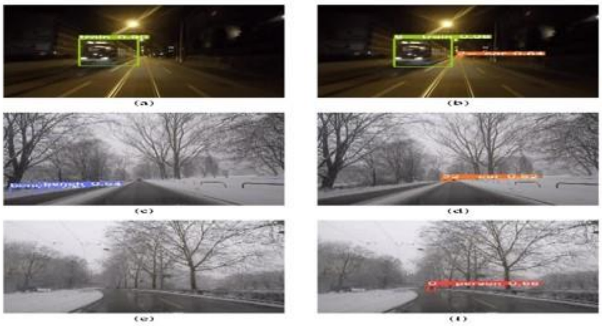


Fig. 2. Detection results on yolov8.

A. Performance evaluation

Important metrics to consider when assessing the effectiveness of the presented models for object detection in pictures and videos include recall, precision, and F1 score. These metrics gave a numerical evaluation of the model's capacity for precise item detection and the avoidance of false positives. The percentage of genuine positives that the model successfully distinguishes from the true positives in the ground truth is known as recall. In other words, it assesses the model's capacity to identify every occurrence of a specific object present in an image. The recall calculation is

$$r = tp / (tp + fn) \tag{1}$$

Contrarily, precision is the proportion of genuine positives among all items the model has classified as positive that have been correctly detected by it. It gauges the model's capacity to steer clear of false positives. For precision, apply the following equation.

$$r = tp / (tp + fp) \tag{2}$$

The F1 score is the harmonic mean of recall and precision, which provides a balanced measure of the model's performance. The formula for the F1 score is

$$f1\ score = 2 * ((p * r) / (p + r)) \tag{3}$$

During training, the objective is to minimize the total loss by adjusting the model parameters, such as the weights and biases, using an optimization algorithm. Monitoring the box loss and class loss separately helped to identify areas where the model may need improvement, such as in localizing objects or correctly classifying them. Similarly, monitoring the val loss helped to identify when the model is overfitting to the training data and may need regularization techniques like dropout or weight decay. mAP@0.5 and mAP@0.95 are in the IoU threshold used for evaluation. mAP@0.5 measures the average precision across IoU thresholds from 0.5 to 1.0, while mAP@0.95 measures the average precision across IoU thresholds from 0.95 to 1.0.

V. CONCLUSION

An innovative autonomous robot has been designed and implemented to collect essential data in hazardous or inaccessible environments. The robot is engineered to navigate and collect data from locations too dangerous or remote for human exploration, unlocking unprecedented possibilities for scientific research and exploration. The AI powered drone, capable of precise human identification through a user-friendly mobile control interface, leverages advanced deep learning approaches and architectures like YOLO for remarkable accuracy in realtime human detection tasks. The robot is equipped with a diverse array of sensors, including cameras, and uses advanced image processing and GPS tracking technologies for thorough data collection and processing. The integration of a notification system for critical situations enhances the robot's responsiveness in challenging environments. This technological amalgamation holds the promise of revolutionizing data collection in previously considered unreachable environments, opening new avenues for scientific discovery, resource assessment, and environmental monitoring. The proposed autonomous robot is a catalyst for transformative advances in scientific exploration, resource evaluation, and environmental surveillance in remote and dangerous locations.

ACKNOWLEDGMENT

The authors gratefully acknowledge the guidance of Dr. Nitha C Velayudhan and service coordinator Mrs. Najla Nazar and Mrs. Gishma KM. Their expertise played an important role in determining the direction and focus of this research.

REFERENCES

[1] Ning Zhang, Francesco Nex, George Vosselman, Norman Kerle, "Training a disaster victim detection network for UAC search and rescue using harmonious composite images", 2022.

[2] Ravindra. R, Patil, Rajnish Kaur Calay, Mohammad. Y, Mustafa, Saniya. M, Ansari, "AI-Driven High-Precision Model for Blockage Detection in Urban Wastewater Systems", 2023.

[3] Zhengxin Zhang, "Article Drone-YOLO: An Efficient Neural Network Method for Target Detection in Drone Images", 2023.

[4] Peng Zhang, Weimin Lei, Xinlei Zhao, Lijia Dong and Zhaonan Lin, "An Adaptive Multi-Scale Network Based on Depth Information for Crowd Counting", 2023.

[5] M.D. Mursalin and Syed Mohammed Shamsul Islam, "Deep Learning for 3D Ear Detection: A Complete Pipeline from Data Generation to Segmentation", 2021.