AI Enabled Robot for Data Collection in Unreachable and Extreme Environment

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

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

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

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

Divya P S
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.**

Keywords*—*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 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 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.

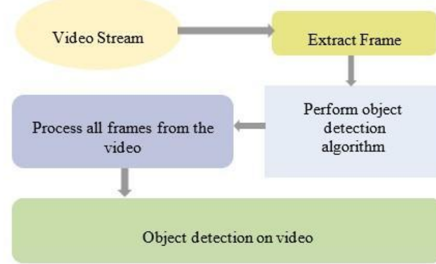

Fig. 1. Process of object detection on video.

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.

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

$$r = tp / (tp + fn) \quad (1)$$

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.

$$r = tp / (tp + fp) \quad (2)$$

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

$$f1\ score = 2 * ((p * r) / (p+ r)) \quad (3)$$

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.