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.

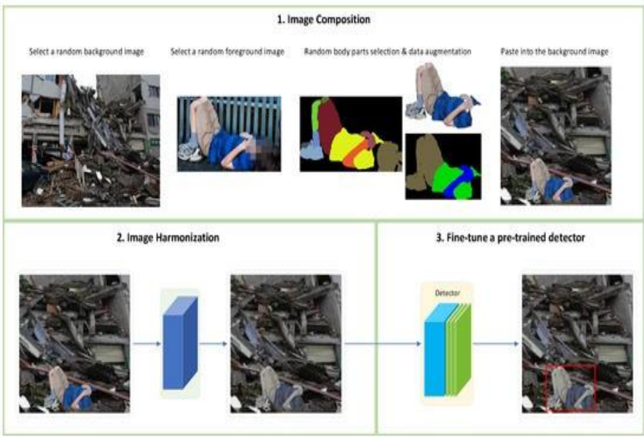

Fig -1: Proposed framework.

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.

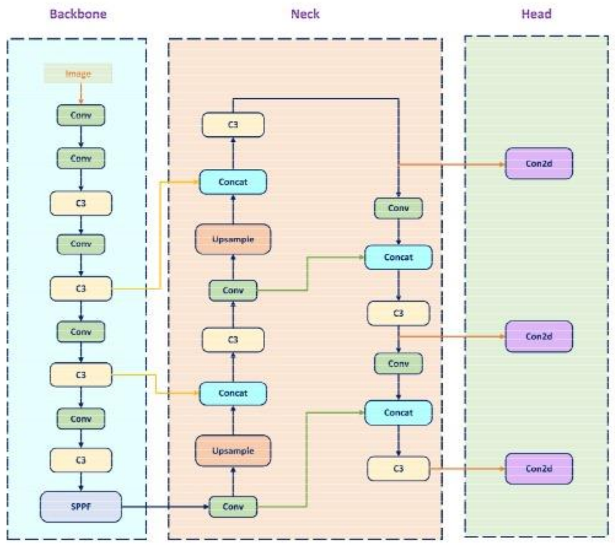

Fig -2: Architectural perception of YOLOv5 model.

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.

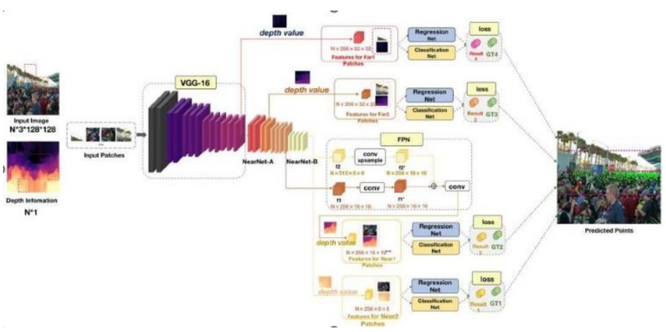

Fig -3: The overall framework of the proposed NF-Net. The backbone of NF-Net adopts all convolutional layers in VGG 16 and a NearNet network.

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.

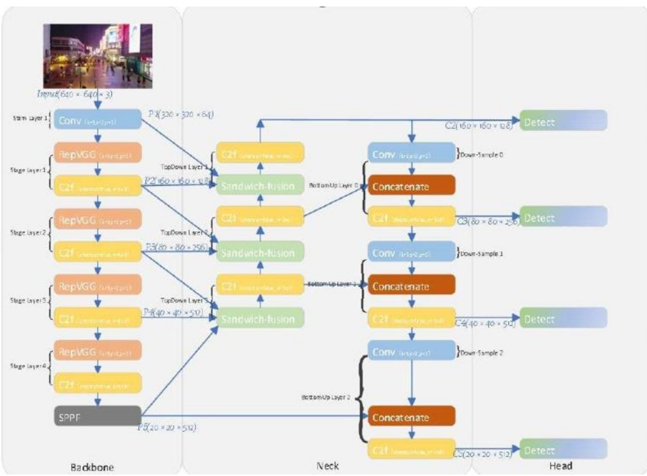

Fig- 4: Proposed network model (large) with its backbone, neck and head.

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.

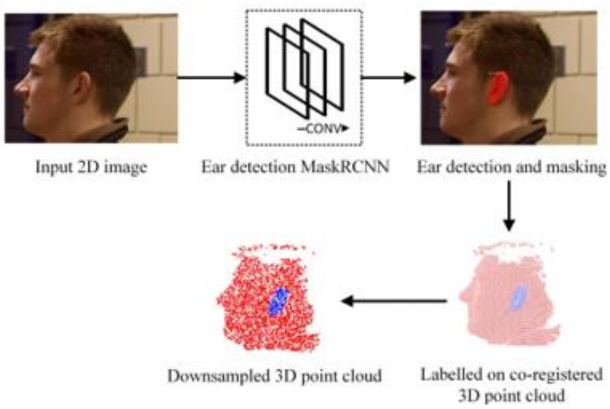

Fig -5: Block diagram of ground-truth labeling procedure on real data.

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.