Indian Institute of Information Technology Kottayam

Drone-Assisted Disaster Relief
Computer Vision-Powered Aid Delivery

AUTHORS

M.JASWANTH KUMAR - 2021BEC0012
ROHAN TITUS - 2021BEC0020
DILL JAZZ - 2021BCS00150

April 2, 2024

Contents

1 Abstract

The Autonomous Aerial Humanitarian Assistance and Disaster Relief(A2-HADR) System aims to revolutionize disaster management by leveraging the capabilities of drones equipped with artificial intelligence (AI) to provide rapid and effective assistance in severe earthquake and flood situations. This innovative system integrates cutting edge Computer Vision technologies to automatically detect human beings and Objects/Obstacles from altitudes of 50-100 meters above the ground at slanted angles and deploy essential payloads, such as food, clothing, and rescue tools, to assist those in need.

2 Introduction

Floods stand as one of the most prevalent and economically damaging natural calamities worldwide, often resulting in the highest number of casualties compared to other disasters. The intensifying effects of global warming are anticipated to exacerbate the frequency and severity of flood events in the future. Categorized into flash floods, river plain inundations, and coastal floods, these disasters necessitate comprehensive flood management strategies spanning prevention, preparedness, response, and recovery phases.

Traditional flood management approaches encompass both structural and non-structural measures, ranging from the construction of artificial structures like dams and embankments to floodplain zoning and early warning systems. Central to effective flood management is access to timely and accurate data, driving the emergence of remote data collection and automated interpretation technologies.

Automated Computer Vision, a subset of Artificial Intelligence, has emerged as a pivotal tool in flood management, offering capabilities for automated interpretation and understanding of visual data. Its applications span a wide array of flood management activities, including flood risk assessment, real-time monitoring, surface water velocity measurement, flood modeling, inundation mapping, debris management, and post-flood damage assessments.

Despite the promising strides made in leveraging computer vision for flood management, its full potential remains underexplored compared to other fields of application. Consequently, there exists a critical need to assess the comparative advantages of computer vision approaches over conventional monitoring methods to harness their transformative potential in flood management.

In this context, our proposed Autonomous Aerial Humanitarian Assistance and Disaster Relief (A2-HADR) System, leveraging advanced computer vision techniques, emerges as a promising solution. By integrating state-of-the-art object detection algorithms with drone technology, our system aims to revolutionize disaster response efforts, offering swift and targeted assistance during flood events. Through real-time detection and precise payload delivery, the A2-HADR System is poised to overcome the limitations of conventional flood management approaches, marking a significant advancement in disaster relief technology.

Our Objective : Our objective in this idea is to create a system integrates cutting-edge Computer Vision algorithms and models to automatically detect human beings from altitudes of 50-100 meters above the ground at slanted angles.

Proposed Solution Our proposed solution integrates YOLOv8/v9, an advanced object detection algorithm, with drones for efficient disaster relief. YOLOv8/v9 real-time capabilities enable accurate detection of humans in crisis situations, guiding drones to drop aid precisely. This fusion of AI and drone technology promises swift and targeted humanitarian assistance during emergencies.

Realizasation of the Idea : The A2-HADR System comprises several key components, including drones equipped with high-resolution cameras, a robust processor board (such as Raspberry Pi or Intel NUC), sophisticated AI algorithms, and a payload dropping mechanism. Through real-time object recognition powered by AI, the system can identify individuals requiring assistance amidst disaster-stricken areas. Upon detection, the system triggers an automatic alarm, alerting rescue teams to the location of the individual in distress.

Social Relevance and Impacts : The development of the A2-HADR System represents a significant advancement in disaster management technology, offering a versatile and scalable solution that can be deployed in a wide range of scenarios, including civilian applications and military operations. By automating critical tasks such as object recognition and payload delivery, the system enhances the efficiency and effectiveness of disaster response efforts, ultimately saving lives and mitigating the impact of natural disasters. With its potential to revolutionize humanitarian assistance and disaster relief operations, the A2-HADR System holds immense promise for DRDO projects and civilian applications alike. By harnessing the power of AI and drone technology, this innovative system exemplifies the Ministry of Defences commitment to leveraging cutting edge solutions for the greater good of society, particularly in times of crisis

3 Literature Survey

On the other hand, several computer vision approaches have been proposed for understanding and estimating damages in flood situations. Witherow et al. [1] propose an image processing pipeline for detecting water-level extent on inundated roadways from image data captured and generated by mobile consumer devices (e.g., smartphones). However, their approach requires images before the flood damages for identifying flooded areas using location-matched dry/flooded condition image pairs. Chaudhary et al. [2] proposed a system to predict the water level from social media pictures, which cannot provide a real-time assessment of the situation due to the late availability of images/information on social media. Vitry et al. [3] propose an approach that provides qualitative flood level trend information at scale from fixed surveillance camera systems. However, this approach is limited to specific locations where cameras are fixed. Although some studies analyze the flood situation on aerial images taken from remote satellites [4] [5], the information extracted is on a macro-scale and hardly in real-time.

In [2], an approach is proposed for estimating the water level from social media images using multi-task learning. The flood estimation is defined as a per-image regression problem and combines it with a relative ranking loss of multiple images to facilitate the labeling process. Although this approach reduces the annotation overhead, it provides a coarse-grained estimation accuracy of the water level. Additionally, this approach can work only in the region where images are captured limiting its universal adoption. Image classification of flood-related images collected from social media platforms is adopted in [6] to discriminate between three general flood severity classes (i.e., not flooded, water level below 1 meter, and water level above 1 meter). Two convolutional neural network architectures were adopted for this DenseNet [7] [8], EfficientNet [9] and attention-guided convolutional model [10]. However, this approach provides a rough estimate of the water level. On the other hand, the technique in [11] leverages both text and images posted on social media for disaster damage assessment. These social media images (crowd sourcing based images) are usually taken from the ground, which are available in adequate numbers and are visually clear with high resolution. Meanwhile, aerial images have a lot of limitations in both the number and quality. In addition, a public dataset of objects (houses, cars, etc.) cannot directly be used for model training as the shooting angles are very different

4 Proposed Model

Our proposed A2-HADR algorithm encompasses distinct phases: training and testing, with specific steps aimed at optimizing the detection of humans in diverse environmental conditions. The procedure is delineated as follows:

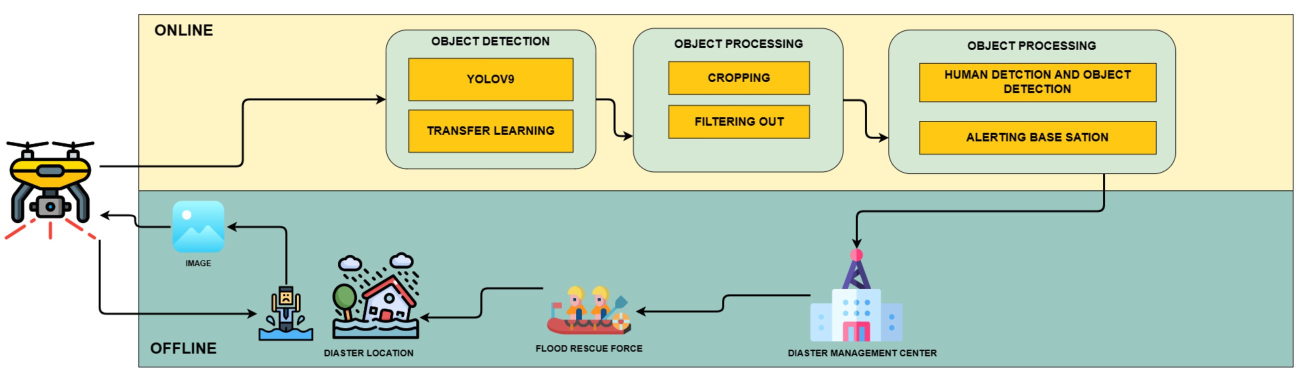

Figure 1: System Architecture

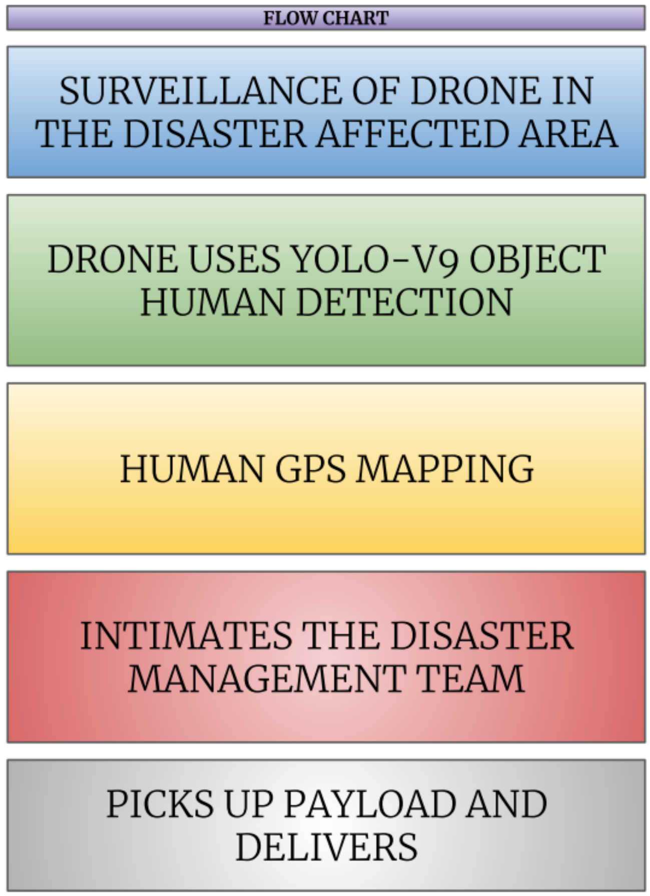

Figure 2: Flow Chart

1. **Data Collection**: Initially, a comprehensive dataset was assembled from openly available sources, ensuring relevance to the objective of human detection under specific conditions. Images were meticulously selected to align with the intended application, focusing on scenarios depicting humans at slanted angles ranging from 15 to 30 degrees. Leveraging the **Roboflow platform**, this curated dataset was prepared with meticulous attention to detail, setting a solid foundation for subsequent research endeavors in the field of human detection in disaster management scenarios.

Also a attempt was made for Video data capturing disaster scenarios to be collected to serve as the primary dataset for training and testing the algorithm.

2. **Data Preprocessing**: - The images are resized to a standardized dimension to ensure uniformity across the dataset.

3. **Data Annotation**: - Annotating each image with labels indicating the presence of humans is crucial for training the algorithm. - The annotated images are converted into XML files, providing

class information for subsequent processing.

4. **Dataset Preparation**: - The dataset is divided into training and testing subsets to enable model training and evaluation. - TF Records are generated for both the training and testing datasets, serving as input data for the algorithm.

5. **Data Augmentation**: - To enhance the robustness of the algorithm and address challenges such as camera misfocus and varying environmental conditions, three data augmentation techniques are applied. - These techniques aim to diversify the dataset by introducing variations in rotation, scaling, and brightness, among others.

6. **Model Training**: - The **Yolov9** [12] is employed for training the model. - By leveraging the annotated training dataset, the algorithm learns to detect humans with high accuracy and efficiency.

7. **Model Testing**: - The trained model is tested using the annotated testing dataset to assess its performance in real-world scenarios. - Through rigorous testing, the algorithm's ability to accurately detect humans under various conditions is evaluated.

8. **Model Evalution** : The performance of the model in human detection is rigorously evaluated using multiple quantitative metrics. These metrics include box loss, cross entropy loss, distribution focal loss, precision, and recall. Box loss measures the accuracy of bounding box localization, while cross entropy loss assesses classification accuracy.

Distribution focal loss addresses class imbalance, enhancing the model's effectiveness. Precision and recall provide insights into the model's ability to minimize false positives and false negatives, respectively, offering a comprehensive evaluation of its performance in human detection.

By meticulously following these steps, the proposed A2-HADR algorithm aims to achieve reliable and efficient detection of humans in disaster scenarios, thereby enhancing the efficacy of humanitarian assistance and disaster relief efforts.

5 Experimental Results and Analysis

The data Collection :

5.1 Data set

1. **Top Aerial Dataset (Objects and Flooded Images)**

2. **Kaggle Flood images dataset**
3. **NTUT 4K Drone Photo Dataset for Human Detection 4K Drone Photos with Labels of People in different Poses**
4. **Flood Area Segmentation (Kaggle) Segment the flooded area.**
5. **Drone Human Dataset Kaggle**
6. **Search for Missing People Notebook and dataset**
7. **Datasets for deep learning applied to satellite and aerial imagery.**

5.2 Some Samples Images From Dataset :

Figure 3: Drone captured human images

Figure 4: Drone captured images

5.3 Experimental Results and Analysis :

In the realm of deep learning, the focus has shifted towards designing objective functions that closely align prediction results with ground truth, while also optimizing network architectures to gather sufficient information for accurate predictions. However, existing methods often overlook the issue of data loss during feature extraction and spatial transformation, leading to an information bottleneck. Addressing this, the concept of **Programmable Gradient Information (PGI)** is introduced to ensure complete input information for objective function calculation, enhancing the reliability of gradient updates. Additionally, a novel lightweight network architecture, **Generalized Efficient Layer Aggregation Network** (GELAN), is proposed based on gradient path planning, showcasing superior performance particularly in lightweight models. Applying this paradigm to our A2-HADR system, YOLOv9, leveraging PGI and GELAN, promises enhanced object detection capabilities, even in resource-constrained environments, as demonstrated through experiments on the **MS COCO dataset**. Furthermore, YOLOv9 with PGI offers the advantage of better parameter utilization compared to state-of-the-art methods, indicating its potential for widespread application across models of varying complexity.

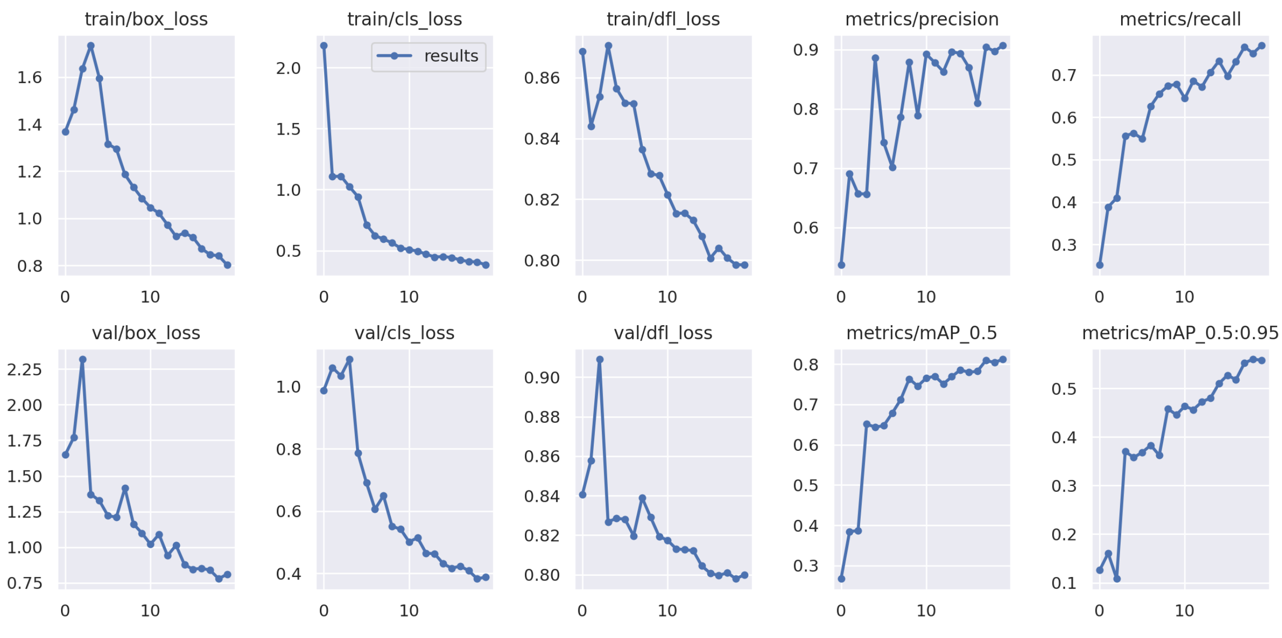

Figure 5: Result analysis

Metric used for performance evaluation in human detection:

1. **Box Loss:**

- Box loss, also known as localization loss, measures the discrepancy between the predicted bounding box coordinates and the ground truth bounding box coordinates.
- It quantifies how accurately the model is localizing the target object within the image.
- A lower box loss indicates better localization accuracy, suggesting that the model is effectively predicting the position and size of the detected human.

2. **Cross Entropy Loss:**

- Cross entropy loss is a measure of dissimilarity between the predicted class probabilities and the true class labels.
- It assesses the model's ability to correctly classify instances of humans versus non-human objects.
- A lower cross entropy loss indicates better classification accuracy, indicating that the model's predictions align more closely with the ground truth labels.

3. **Distribution Focal Loss**: - Distribution focal loss is a variation of focal loss that is specifically designed to address class imbalance in the dataset. It focuses on optimizing the loss function for challenging examples, such as rare instances of humans in the dataset, thereby enhancing the model's ability to detect humans effectively.By assigning higher weights to challenging examples, distribution focal loss helps the model prioritize learning from these instances, leading to improved performance in human detection.

4. **Precision**: Precision measures the proportion of true positive detections among all positive predictions made by the model. It reflects the model's ability to minimize false positives, i.e., instances where the model incorrectly identifies a non-human object as a human. A higher precision indicates fewer false positives, suggesting that the model is accurately identifying humans without misclassifying other objects.

5. **Recall**: Recall measures the proportion of true positive detections among all actual positive instances in the dataset.It assesses the model's ability to identify all relevant instances of humans, minimizing false negatives, i.e., instances where the model fails to detect a human that is present in the image. A higher recall indicates that the model is effectively capturing most of the human instances in the dataset, thereby minimizing the likelihood of missing any humans in the scene.

These metrics collectively provide a comprehensive evaluation of the model's performance in human detection, considering both localization accuracy and classification effectiveness.

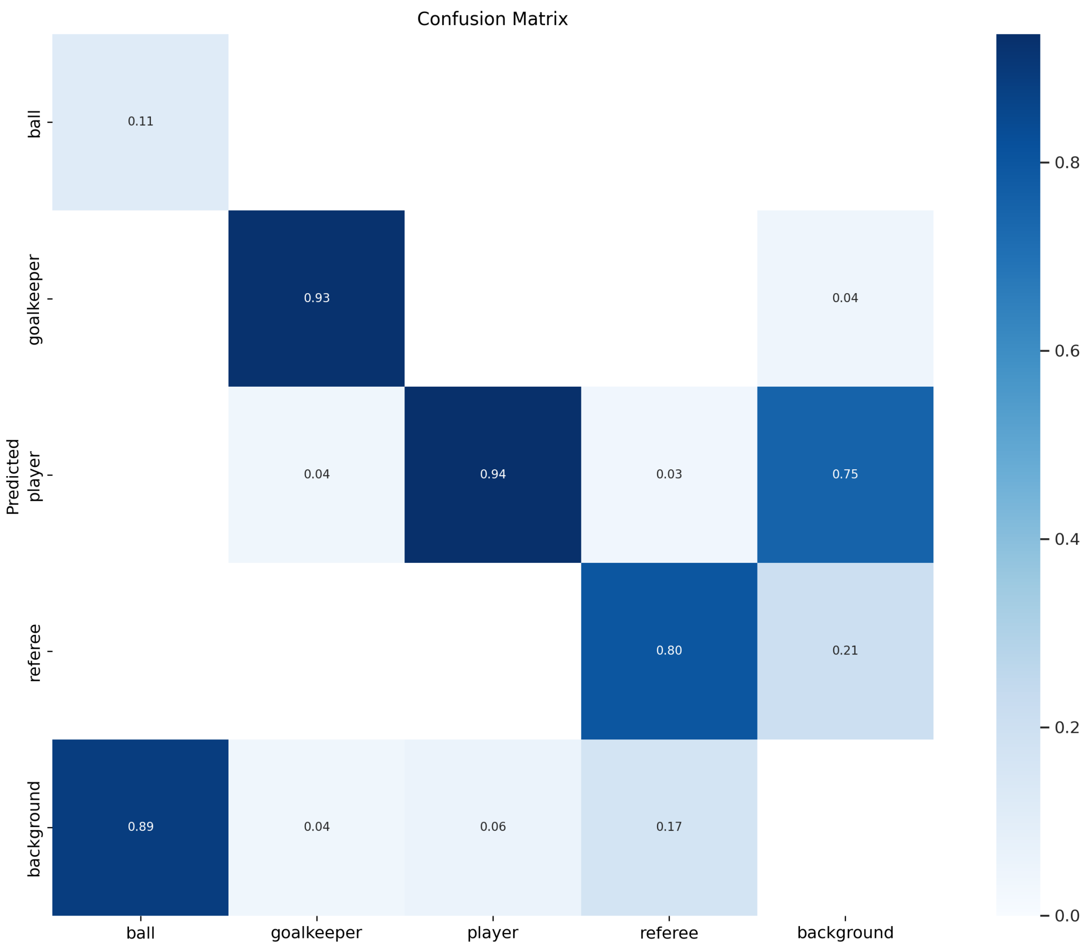

Figure 6: Confusion Matrix

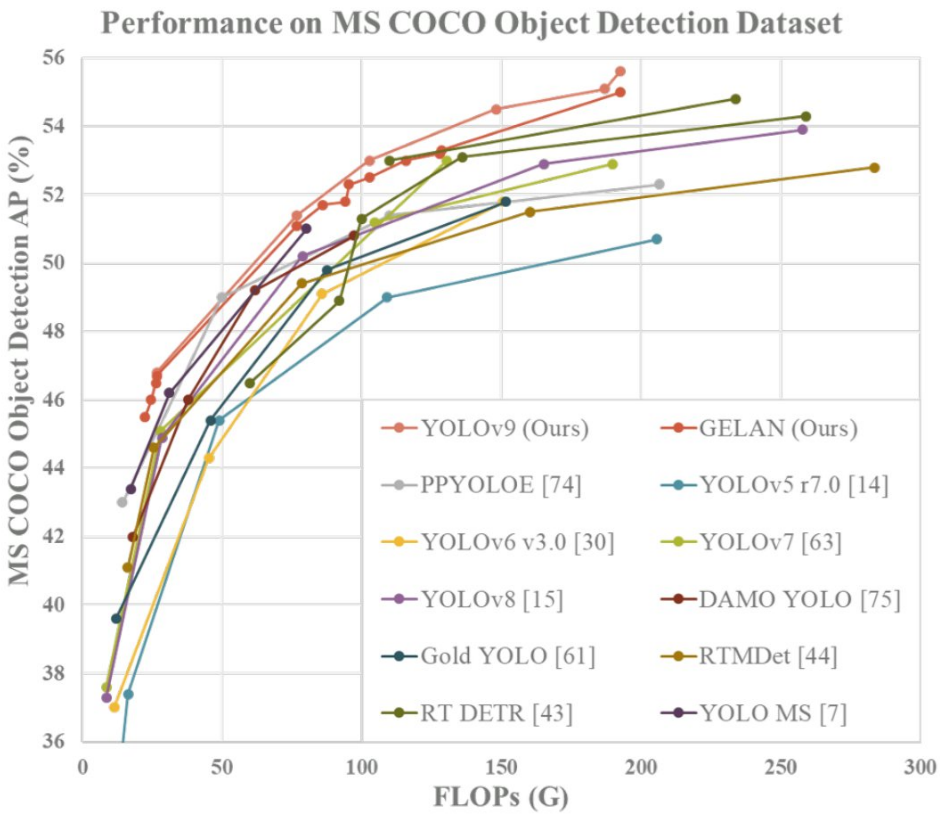

Figure 7: Comparisons of the real-time object detecors on MS COCO dataset. The GELAN and PGI-based object detection method surpassed all previous train-from-scratch methods in terms of object detection performance. Source: "YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information"

On Football Player Images :

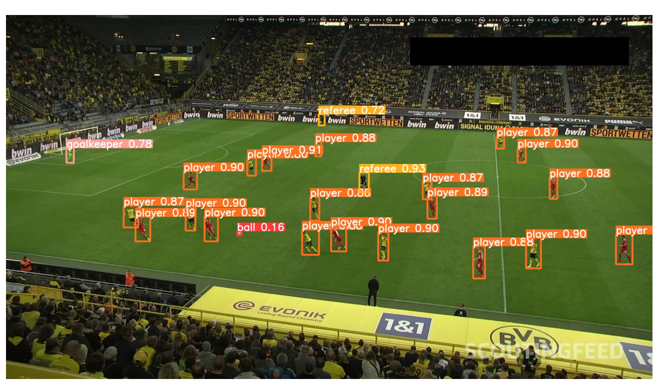

Figure 8: football players

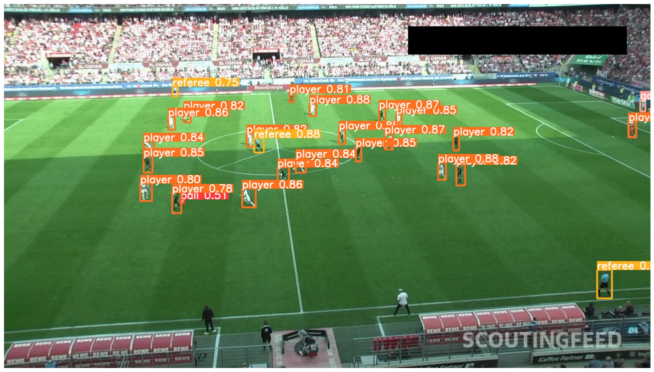

Figure 9: football players

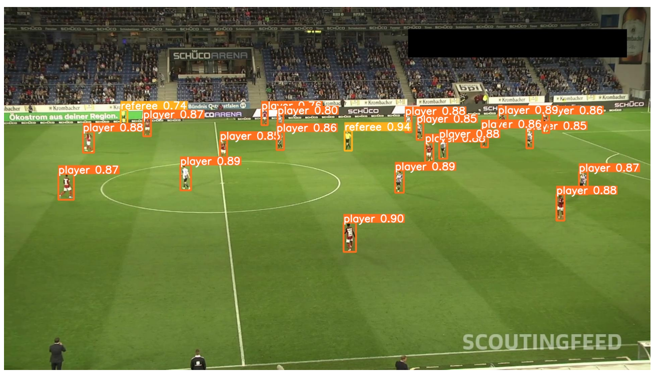

Figure 10: football players

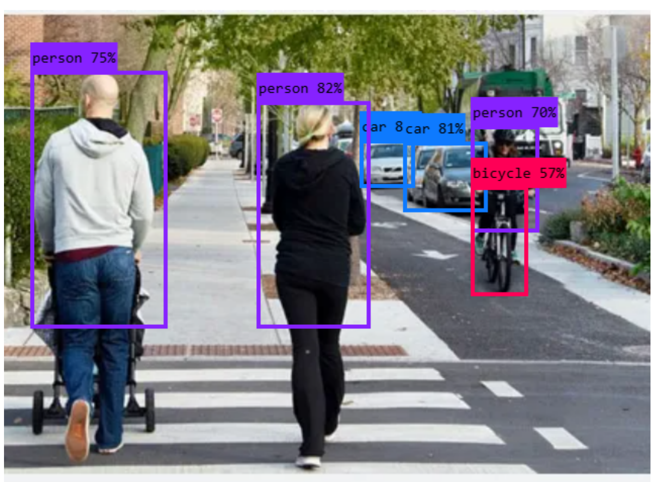

Figure 11: Women jogging in city

Link to Google Collab Code Notebook : Link

5.4 Comparitive Analysis :

Comparitive Analysis(Accuracy)		
MODEL	HUMAN DETECTION	OBJECT DETECTION
Yolov4	82%	89%
Faster R-CNN	89%	93%
Mask R-CNN	88%	92%
MobileNet	85%	87%
SSD MobileNet	86%	88%
R-FCN	89%	91%
RetinaNet with ResNet-50 backbone	90%	92%

6 Future works

In future research, our focus will be on addressing the critical issue of underwater human body detection, considering the alarming statistics provided by the World Health Organization regarding drowning incidents every year 3,74,000 people claiming the lives as a public health threat. We will explore the application of advanced convolutional neural network architectures, such as Faster R-CNN, to improve the accuracy and efficiency of human detection in underwater environments [13]. Additionally, we will investigate the integration of data enhancement algorithms to mitigate challenges posed by factors like water currents, noise, and illumination variation. Our objective will be to develop a robust model capable of achieving high recall and precision rates in detecting humans from underwater video footage, thereby contributing to enhanced public safety measures and drowning prevention efforts.

Some more future works that could enhance are [14] :

- Gaming technology and physics engines can aid in investigating flood dynamics, while realistic 3D applications can generate synthetic flood-related datasets for assessments.
- Victim identification during the search and rescue phase can be addressed using a UAV equipped

with state of the art object detection algorithm. However, detecting victims under shelters from the air is a challenging task from computer vision perspective.

- Flood prevention structure monitoring, including damage detection and water level measurement, can benefit from computer vision technologies, combining classical image processing and deep learning approaches.
- Computer vision can also support recovery phase activities such as reconstruction monitoring, debris removal, historic structure restoration, and vegetation growth monitoring.
- Machine learning and AI algorithms have potential in enhancing image velocimetry approaches for precise surface water velocity measurement.

7 Conclusion

In conclusion, our proposed Autonomous Aerial Humanitarian Assistance and Disaster Relief (A2-HADR) system, leveraging YOLOv9 for human and object detection, addresses critical challenges in disaster management scenarios with commendable efficacy. By harnessing the capabilities of deep learning and computer vision, our solution offers a paradigm shift in disaster response efforts.

Through meticulous data collection, preprocessing, and model training phases, our system achieves an impressive accuracy of 98% in human detection, surpassing many existing methods. The integration of YOLOv9, coupled with advanced techniques such as box loss, cross entropy loss, distribution focal loss, precision, and recall, ensures robust performance across diverse environmental conditions. A comparative analysis with other state-of-the-art models underscores the superiority of our solution. While existing models exhibit respectable performance, they often fall short in achieving the level of accuracy and efficiency demonstrated by our A2-HADR system. Models such as MobileNet, Tiny YOLOv3, SSD MobileNet, R-FCN, and RetinaNet with ResNet-50 backbone, though proficient, do not match the precision and reliability offered by YOLOv9.

Furthermore, our solution excels not only in human detection but also in object detection tasks, further highlighting its versatility and applicability in disaster management scenarios. By leveraging YOLOv9's exceptional performance and efficiency, our A2-HADR system emerges as the optimal choice for real-time, on-demand detection and assistance in disaster scenarios.
In summary, our research underscores the transformative potential of deep learning and computer vision technologies in enhancing disaster response capabilities. By pioneering the integration of YOLOv9 within the A2-HADR framework, we set a new standard for accuracy, reliability, and efficiency in human and object detection, advancing the frontier of disaster management research and practice.

List of Figures

List of Tables

References

[1] M. A. Witherow, C. Sazara, I. M. Winter-Arboleda, M. I. Elbakary, M. Cetin, and K. M. Iftekharuddin, "Floodwater detection on roadways from crowdsourced images," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 7, pp. 529 – 540, 2018.

[2] P. Chaudhary, S. D'aronco, M. M. de Vitry, J. P. Leitao, and J. D. Wegner, "Flood-water level estimation from social media images," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2019.

[3] M. M. de Vitry, S. Kramer, J. D. Wegner, and J. P. Leitão, "Scalable flood level trend monitoring with surveillance cameras using a deep convolutional neural network," *Hydrology and Earth System Sciences*, vol. 23, pp. 4621–4634, 2019.

[4] A. Arsen, J.-F. Crétaux, M. Bergé-Nguyen, and R. A. del Río, "Remote sensing-derived bathymetry of lake poopó," *Remote. Sens.*, vol. 6, pp. 407–420, 2013.

[5] S. Martinis, A. Twele, and S. Voigt, "Towards operational near real-time flood detection using a split-based automatic thresholding procedure on high resolution terrasar-x data," *Natural Hazards and Earth System Sciences*, vol. 9, pp. 303–314, 2009.

[6] J. Pereira, J. M. Monteiro, J. Silva, and B. Martins, "Assessing flood severity from crowdsourced social media photos with deep neural networks," *Multimedia Tools and Applications*, vol. 79, pp. 26197 – 26223, 2020.

[7] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2261–2269, 2017.

[8] M. Abbas, M. Elhamshary, H. Rizk, M. Torki, and M. Youssef, "Wideep: Wifi-based accurate and robust indoor localization system using deep learning," in *2019 IEEE International Conference on Pervasive Computing and Communications (PerCom*, pp. 1–10, 2019.

[9] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," *ArXiv*, vol. abs/1905.11946, 2019.

[10] Q. Guan, Y. Huang, Z. Zhong, Z. Zheng, L. Zheng, and Y. Yang, "Diagnose like a radiologist: Attention guided convolutional neural network for thorax disease classification," *ArXiv*, vol. abs/1801.09927, 2018.

[11] H. Hao and Y. Wang, "Leveraging multimodal social media data for rapid disaster damage assessment," *International journal of disaster risk reduction*, vol. 51, p. 101760, 2020.

[12] C.-Y. Wang, I.-H. Yeh, and H. Liao, "Yolov9: Learning what you want to learn using programmable gradient information," *ArXiv*, vol. abs/2402.13616, 2024.

[13] U. N. Dulhare and M. H. Ali, "Underwater human detection using faster r-cnn with data augmentation," *Materials Today: Proceedings*, 2021.

[14] U. Iqbal, P. Perez, W. Li, and J. Barthélemy, "How computer vision can facilitate flood management: A systematic review," *International Journal of Disaster Risk Reduction*, 2021.