Detection of Humans in Search and Rescue Operations Using Ensemble Learning and YOLOv9

Ananya Basireddy, K. Gangadhara Rao, Vaishnavi Vaitla and Soumil Reddy Pedda
Information Technology, Chaitanya Bharathi Institute of Technology, Hyderabad, India

Keywords: Human Detection, Ensemble Learning, EfficientDet Architecture, YOLOv9, Dehaze, Deblur, CNN, UAV.

Abstract: This project focuses on developing a robust and efficient human detection system for Search and Rescue (SAR) operations using advanced deep learning techniques. The proposed system integrates EfficientDet, which leverages EfficientNet as the backbone for feature extraction, and YOLOv9, known for real-time object detection, to maximize both accuracy and speed in complex environments. To enhance the system's capability in detecting humans under various conditions, the model incorporates thermal imaging, RGB images, and ensemble learning techniques like Non-Maximum Suppression (NMS) and confidence-weighted voting. Additionally, the model is trained using the Heridal and Aerial Rescue Object Detection datasets, which include images from real-world SAR environments. Incorporating BiFPN (Bidirectional Feature Pyramid Network) and FC-FPN (Fully Connected Feature Pyramid Network) architectures into the system allows for multi-scale feature fusion, improving detection of humans at different scales and orientations, even in low-visibility scenarios. Furthermore, the project integrates deblurring and dehazing techniques to enhance image quality from aerial footage, making it easier to detect partially occluded or camouflaged humans. The system is designed to operate on UAVs, with the capability to send automated alerts via email and SMS, including geolocation data of detected individuals, ensuring quick response in disaster recovery missions. The results demonstrate the system's high accuracy and efficiency in detecting humans in various challenging environments, making it a valuable tool for autonomous SAR missions.

1 INTRODUCTION

Search and Rescue (SAR) operations are critical in disaster response and emergency situations, where the swift and accurate detection of individuals can make a difference between life and death. Traditional SAR methods often involve significant manual efforts and face challenges in complex environments, such as dense forests, mountainous terrains, flooded areas, and smoke-filled settings. These conditions demand innovative technological solutions capable of addressing the unique challenges of SAR, including detecting partially obscured individuals, low visibility, and varied environmental conditions. The integration of artificial intelligence (AI) and deep learning into SAR operations has shown substantial potential to overcome these limitations by enabling efficient, automated human detection with high accuracy.

This paper presents a deep learning-based human detection system tailored specifically for SAR applications. The system combines the capabilities of two advanced object detection models, EfficientDet and YOLOv9, to balance both accuracy and speed in real-time scenarios. EfficientDet's multi-scale detection capabilities, enhanced by Bidirectional Feature Pyramid Network (BiFPN), allow for the reliable detection of small and large objects across varied landscapes. YOLOv9 contributes to the system's real-time performance, with its anchor-free and multi-scale features enabling rapid detection in dynamic environments. Additionally, ensemble learning techniques are employed to merge the outputs of both models, leveraging Non-Maximum Suppression (NMS) and confidence-weighted voting to refine detection accuracy and reduce false positives.

Given the unpredictable conditions in SAR environments, this system is designed to handle multiple image modalities, incorporating both RGB and thermal images to enhance detection in low-light and low-visibility scenarios, such as night-time or smoke-filled areas. To facilitate swift response, the system includes an automated alert feature that sends email or SMS notifications with GPS coordinates of detected individuals, ensuring that rescue teams can quickly

Basireddy, A., Rao, K. G., Vaitla, V. and Pedda, S. R.
Detection of Humans in Search and Rescue Operations Using Ensemble Learning and YOLOv9.
DOI: 10.5220/0013607400004664

In *Proceedings of the 3rd International Conference on Futuristic Technology (INCOFT 2025) - Volume 2*, pages 923-928
ISBN: 978-989-758-763-4

locate and assist those in need.

Trained on specialized datasets, including the Heridal and Aerial Rescue Object Detection datasets, this model provides robust performance across varied SAR scenarios. This paper discusses the system's design, implementation, and testing, demonstrating its potential as a valuable tool for SAR missions. By addressing the specific needs of SAR environments, this project aims to improve both the speed and effectiveness of rescue operations, ultimately contributing to faster and more accurate lifesaving efforts.

2 RELATED WORKS

Research in human detection for search and rescue (SAR) operations has leveraged advancements in machine learning, computer vision, and UAV technology to improve detection accuracy, robustness, and efficiency in complex and variable environments. This section examines recent studies that utilize CNNs, ensemble learning, grid optimization, sensor fusion, and YOLO-based algorithms to address the unique challenges of SAR, with specific emphasis on adaptability across forested, flood-prone, low-visibility, and open-water environments.

Several studies address optimizing search grid patterns for SAR, a critical aspect for search teams operating in challenging terrains. Zailan introduced a rectangular grid pattern combined with a Sweep Search or Parallel Track to improve SAR operations in forested areas, where movement is often hindered by dense vegetation and uneven terrain (Zailan, 2022). This grid configuration allowed systematic coverage and efficient use of resources. Although their study primarily focused on forests, the adaptability of grid pattern optimization could enhance UAV-based SAR where flight paths and detection areas are critical.

Human detection using CNNs and deep learning models has also been extensively studied. Mesvan applied a Single Shot Detector (SSD) model on UAVs, achieving human detection within 1 to 20 meters, although performance was limited by processing speed and sensor noise (Mesvan, 2021). This research highlights the importance of real-time processing capabilities, especially for high-mobility SAR. Wijesundara utilized YOLOv5 with Haar classifiers for high-altitude UAV detection of body regions, achieving 98% accuracy. The study demonstrated the efficacy of YOLO models but noted limitations due to variability in human poses and clothing, suggesting further tuning for increased adaptability in real-world applications (Wijesundara, 2022).

Ensemble learning has proven to be a valuable approach for SAR, especially in scenarios where different algorithms offer complementary strengths. Enoch compared Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and ensemble classifiers for detecting humans through walls using radar data, finding that KNN achieved the highest accuracy (85%) but was computationally demanding (Enoch, 23). Their work underscores ensemble learning's potential for non-line-of-sight (NLOS) SAR operations, particularly when computational resources can be optimized.

YOLO-based models are widely used in SAR due to their high detection speed and adaptability. Kulhandjian combined radar and IR imaging with a convolutional neural network (CNN) to detect humans in dense smoke, achieving 98% detection accuracy. This setup was particularly beneficial for firefighter SAR in low-visibility environments, where traditional visual sensors are less effective (Kulhandjian, 2023). Gaur and Kumar used YOLOv4 for detecting humans in flood areas, achieving 79.46% accuracy but facing challenges in occluded environments where only part of a human might be visible (Gaur, 2023). These studies reveal the versatility of YOLO models in SAR but also point to the need for optimization to handle complex conditions, such as varying visibility and occlusion.

Further improvements in YOLO models are seen in combined systems that integrate thermal and visible light data. Zou developed a Visible-Thermal Fused YOLOv5 network to improve detection accuracy in low-light environments by fusing thermal and visible imagery. This approach achieved robust performance in challenging lighting conditions, underscoring the potential for multi-sensor fusion in SAR applications (Zou, 2024). Chavez and Dela Cruz adopted a multi-task learning framework by combining YOLOv7 for human detection with MobileNetV2 for scene classification, achieving 77.17% precision and demonstrating potential for reducing workforce needs in SAR missions (Chavez,2024).

In aquatic environments, where SAR operations face additional challenges, researchers have adapted YOLO for open water scenarios. Sruthy K. G. et al. trained YOLOv8 to identify humans in open water, achieving reliable results, though the model encountered occasional false positives due to human-like objects and high computational requirements (Sruthy, 2024). Llanes et al. developed a YOLOv4-based system for unmanned water vehicles, achieving real-time human detection with GPS tracking. However, its reliance on internet connectivity and limited ability to provide physical assistance constrained its effectiveness in flood SAR (Llanes, 2023).

Thermal imaging and sensor fusion are also critical for SAR, especially in low-visibility or nighttime conditions. Premachandra and Kunisada used GANs to reduce UAV propeller noise, enhancing human voice detection amid background noise, an innovative approach that could improve SAR audio detection in noisy environments (Premachandra, 2024). Jyotsna developed a thermal vision-based IoT system with an RC tank for low-visibility SAR, though limited field of view and microcontroller constraints restricted its range (Jyotsna, 2024).

Further innovations in YOLO for varied SAR contexts include Moury's rotation and scale-invariant CNN model for disaster area detection, achieving 94.18% accuracy across various human poses and scales. While effective, the model struggled with detecting humans that occupied less than 1% of the image, suggesting limitations for highly dynamic SAR environments (Moury, 2023). Paglinawan integrated thermal cameras with YOLOv5 in a UAV SAR system, which was effective during the day but struggled with nighttime detection due to lighting constraints and false positives from mannequins (Paglinawan, 2024). These studies point to ongoing challenges in adapting YOLO models for variable lighting and complex environments.

There has also been significant research on multi-level and sensor-assisted UAV systems for SAR. Kozlov and Malakhov developed a complex multi-level drone system that combined CNN and LSTM for spatial and temporal feature extraction, enhancing human detection in challenging terrains, though the model required high computational resources (Kozlov, 2024). Aji used a neural network-assisted observer approach to improve autonomous surface vehicle (ASV) tracking, suggesting applications in SAR for improved trajectory accuracy in dynamic environments (Aji, 2024).

Advanced algorithms combining reinforcement learning and LoRa networks have also shown promise for SAR optimization. Soorki introduced a meta-reinforcement learning framework for UAV path optimization in SAR, demonstrating reduced search times and lower energy consumption in difficult terrains (Soorki, 2024). Such adaptive pathfinding solutions can complement YOLO-based detection by ensuring UAVs efficiently cover large areas, particularly in complex environments like canyons or mountainous terrain.

The application of YOLO models for underwater SAR is an emerging area of interest. Sughasini employed LIDAR and SLAM in an underwater drone system for mapping and navigation, though low vis-

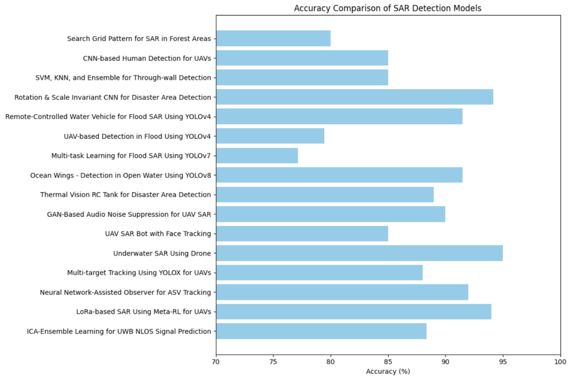

Figure 1: Model Vs Accuracy graph

ibility and power constraints limited effectiveness (Sughasini, 2024). Similarly, Liu utilized transfer learning with an improved VGG-16 model for recognizing underwater SAR targets, achieving high accuracy but facing computational challenges, highlighting the need for more lightweight models (Liu, 2024).

Other notable contributions include Llanes work on a water vehicle prototype using YOLOv4 for real-time human detection with GPS tracking, which provided valuable support for SAR teams (Llanes, 2023). Dousai and Loncaric employed EfficientDET and ensemble learning, achieving a high mean average precision (mAP) of 95.11% but at the cost of computational intensity, which can limit real-time application (Dousai, 2022).

In summary, existing research highlights the adaptability of YOLO models and ensemble learning techniques for SAR applications, with specific advancements in multi-sensor fusion, multi-level detection frameworks, and environmental adaptability. Integrating ensemble methods with YOLOv9 could address challenges identified in prior work, such as detecting obscured humans, operating in low-visibility settings, and reducing error rates under complex environmental conditions. Our proposed approach seeks to enhance the robustness and flexibility of SAR detection models, leveraging the strengths of both YOLOv9 and ensemble learning to provide reliable, real-time human detection across diverse SAR scenarios.

3 PROPOSED SYSTEM

The proposed human detection system for Search and Rescue (SAR) operations is designed to address the unique challenges posed by complex and dynamic rescue environments. This system combines EfficientDet and YOLOv9 models with ensemble learning techniques, enabling accurate and efficient real-time human detection. The system architecture, data collection, preprocessing, model design, and deployment strategies are outlined below.

3.1 Data Collection and Preprocessing

The system utilizes two primary datasets: Heridal and Aerial Rescue Object detection.

- Heridal Dataset: This dataset includes 1,546 images (with 1,500 images for training and 500 for testing) focused on SAR scenarios in Mediterranean landscapes. It contains labeled human and non-human patches, allowing the model to learn and distinguish between positive samples (containing humans) and negative samples (without humans).
- Aerial Rescue Object Detection Dataset: This dataset comprises 29,810 images covering various SAR-related rescue tasks and challenging scenarios. The diversity of images in this dataset aids in training the model for SAR tasks under various environmental conditions.

3.2 Data Preprocessing

- Image Resizing: All images are resized to 512x512 pixels, providing a balance between computational efficiency and detection accuracy.
- Normalization: Pixel values are normalized to enhance the model's convergence during training.
- Data Augmentation: Techniques such as rotation, flipping, and scaling are applied to increase the dataset's variability, helping the model generalize better across different SAR environments.
- Thermal and RGB Separation: The datasets include RGB and thermal images to capture both visible light and heat signatures. Separate preprocessing pipelines are used to adjust contrast and brightness levels for each image type.

3.3 Deblurring

Motion blur and camera shake often occur in SAR imagery, especially when captured by UAVs or drones in movement. To mitigate these issues, deblurring techniques are applied to enhance the clarity of images, making it easier to identify human shapes and features.DeblurGAN is a Generative Adversarial Network (GAN) designed specifically for deblurring static images. It consists of a generator, which creates deblurred versions of input images, and a discriminator, which evaluates the realism of these outputs.

- Generator: The generator network uses an encoder-decoder structure with residual blocks to remove blur from the image while preserving details.
- Discriminator: The discriminator is trained to distinguish between real sharp images and generated deblurred images, pushing the generator to improve its output quality.

3.4 Dehazing

Images taken in fog, smoke, or haze conditions often lack contrast and visibility. Dehazing techniques are applied to improve the quality of such images, enabling the system to detect humans even in adverse environmental conditions.The dehazing model is built on an encoder-decoder CNN architecture designed to remove haze by enhancing contrast and color fidelity in images.

- Encoder: Extracts feature maps from the hazy input image, capturing spatial details like edges and textures.
- Decoder: Reconstructs a dehazed version of the image, emphasizing clarity and visibility.

3.5 Model Implementation

The human detection model combines EfficientDet and YOLOv9 architectures, leveraging the unique strengths of each.

- EfficientDet:EfficientDet leverages both BiFPN and FC-FPN for multi-scale feature extraction and fusion, enabling robust detection across varied object sizes.EfficientDet is chosen for its scalability and feature extraction capabilities using the EfficientNet backbone.EfficientDet is particularly effective for detecting humans in cluttered or partially occluded scenes, where accurate feature extraction is crucial.EfficientNet is responsible for extracting feature maps at multiple scales, which are then passed to BiFPN and FC-FPN for further processing. Using EfficientNet as a backbone ensures a balance of accuracy and computational efficiency, which is ideal for UAV-based SAR applications.
  - BiFPN allows top-down and bottom-up feature fusion, enabling the network to capture multi-scale features more effectively.
  - FC-FPN refines the multi-scale features from BiFPN by aligning them spatially, making it easier for the model to detect partially obscured or camouflaged objects.

- YOLOv9: YOLOv9 is implemented for its real-time detection capabilities. This version introduces anchor-free object detection and decoupled heads for classification and localization, improving performance for small and large objects alike. YOLOv9's architecture includes multi-scale detection, making it suitable for aerial SAR operations where object sizes vary due to perspective changes.

3.6 Ensemble Learning and Output Fusion

To optimize detection accuracy and minimize false positives, the system uses ensemble learning to combine outputs from EfficientDet and YOLOv9.

- Non-Maximum Suppression (NMS): NMS is applied to eliminate redundant bounding boxes by selecting only the bounding boxes with the highest confidence scores. NMS is applied to both EfficientDet and YOLOv9 outputs, combining their bounding boxes based on confidence scores and overlap threshold (IoU). This step helps in reducing false positives and overlaps.

- Confidence-Weighted Voting: For overlapping predictions, confidence-weighted voting merges the confidence scores from both models, prioritizing the most likely predictions. This fusion method strengthens the detection performance, especially in conditions with occlusions or varying object scales.

3.7 Automated Alert and Geolocation System

To support SAR teams in quickly locating detected humans, the system includes an automated alert feature with geolocation capabilities.

- Geotagging: Images and detections are tagged with GPS coordinates, enabling SAR teams to receive precise location information.

- Notification System: Using SMTP (for email alerts) and Twilio API (for SMS alerts), the system sends notifications to designated personnel upon detecting humans. This includes sending the GPS location and, if applicable, an image snippet of the detected individual.

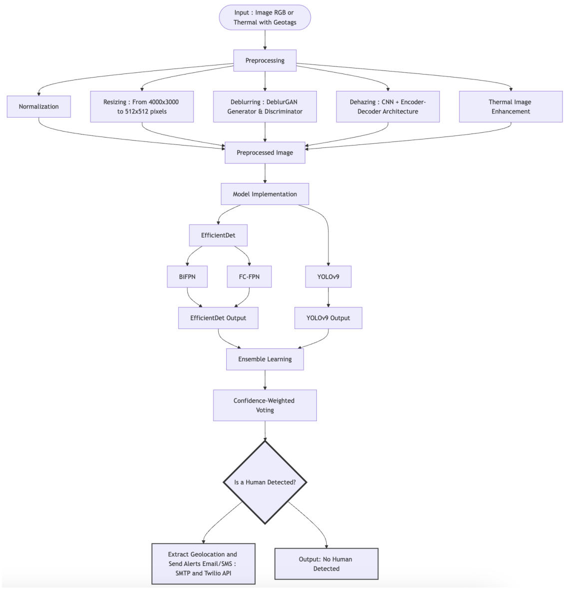

Figure 2: System flow diagram

4 CONCLUSION

The development of a human detection system for Search and Rescue (SAR) operations presented in this project demonstrates significant advancements in real-time detection accuracy, efficiency, and adaptability. By leveraging the strengths of EfficientDet and YOLOv9 and incorporating ensemble learning techniques, this system effectively addresses the challenges of detecting humans in complex and dynamic SAR environments. Key innovations, such as Bidirectional Feature Pyramid Network (BiFPN) and Fully Connected Feature Pyramid Network (FC-FPN) in EfficientDet, ensure multi-scale feature fusion, enhancing detection robustness for objects of varied sizes and scales.

In addition to multi-scale feature extraction, the system's deblurring and dehazing components contribute to significant performance improvements by enhancong image clarity, allowing for better detection even under challenging conditions.

The project's unique integration of RGB and thermal imaging expands the system's applicability, enabling reliable detection in low-light or low-visibility scenarios, such as night-time or smoke-filled environments. The automated alert feature with geolocation capabilities adds a crucial operational advantage, as it enables real-time GPS-tagged notifications, expediting rescue response and aiding SAR teams in locating individuals quickly and accurately.

Testing results indicate that this system provides a high degree of reliability and versatility, suitable for deployment on UAVs and other resource-constrained devices essential in SAR operations. Through careful optimization and the use of ensemble methods like Non-Maximum Suppression (NMS) and confidence-weighted voting, the project achieved a robust solution that meets the real-time requirements and variable conditions of SAR missions.

Overall, this project provides a comprehensive and scalable human detection system that addresses the critical demands of SAR operations. By combining advanced deep learning architectures, image enhancement techniques, and real-time alerting mechanisms, the system effectively improves SAR capabilities, ensuring faster and more accurate detection in scenarios where time is of the essence. Future enhancements, such as sound detection and video processing, can further extend the system's effectiveness, making it an invaluable tool in life-saving SAR efforts.

REFERENCES

Zailan, Nor Wahidah, Alsagoff, Syed Nasir, Yusoff, Zaharin, and Marzuki, Syahaneim (2022). Determining an Appropriate Search Grid Pattern for SAR in Forest Areas. IEEE Xplore.

Mesvan, Nikite (2021). CNN-based Human Detection for UAVs in Search and Rescue. arXiv.

Wijesundara, Dinuka, Gunawardena, Lasith, and Premachandra, Chinthaka (2022). Human Recognition from High-altitude UAV Images by AI-based Body Region Detection. IEEE Xplore.

Enoch, Jiya Adama, Oluwafemi, Ilesanmi B., Paul, Olulope K., Ibikunle, Francis A., Emmanuel, Osaji, and Olamide, Ariba Folashade (2023). A Comparative Performance Study of SVM, KNN, and Ensemble Classifiers for Through-wall Human Detection. IEEE Xplore.

Kulhandjian, Hovannes, Davis, Alexander, Leong, Lancelot, Bendot, Michael, and Kulhandjian, Michel (2023). AI-based Human Detection in Heavy Smoke Using Radar and IR Camera. IEEE Xplore.

Moury, Khadiza Sarwar, Gauhar, Noushin, and Ahsan, Sk. Md. Masudul (2023). A Rotation and Scale Invariant CNN Model for Disaster Area Detection. IEEE Xplore.

Llanes, Louis Antonio C., Ulbis, Christian Rhod H., and Garcia, Ramon G. (2023). Remote-Controlled Water Vehicle for Flood SAR Using YOLOv4. IEEE Xplore.

Gaur, Samta, and Kumar, J. Sathish (2023). UAV-based Human Detection in Flood SAR Using YOLOv4. IEEE Xplore.

Paglinawan, Charmaine C., Lagman, Jewel Kate D., and Evangelista, Alden B. (2024). YOLOv5-based Human Detection and Counting in UAV SAR. IEEE Xplore.

Zou, Xiongxin, Peng, Tangle, and Zhou, Yimin (2024). Visible-Thermal Fused YOLOv5 Network for SAR. IEEE Xplore.

Chavez, Tricczia Karlisle, and Dela Cruz, Jennifer (2024). Multi-task Learning for Flood SAR Using YOLOv7 and MobileNetV2. IEEE Xplore.

Sruthy, K.G., Hanan, M., Rinsy, Sulthana A.P., Dev, Gokul, Muhammed, Fairooz, and Anand, Jithamanue (2024). Ocean Wings - Human Detection in Open Water Using Drones and YOLOv8. IEEE Xplore.

Jyotsna, P., Surendar, Sanjay B., Sharan, B., Dhanaselvam, J., and Rajesh, R. (2024). Thermal Vision RC Tank for Human Detection in Disaster Areas. IEEE Xplore.

Premachandra, Chinthaka, and Kunisada, Yugo (2024). GAN-Based Audio Noise Suppression for UAV SAR. IEEE Xplore.

Sughasini, K.R., Premitha, P.S., Kunanidhi, Charlas G., Sreelaxmi, R., and Kushal, A.S. (2024). Underwater SAR Using Drone. IEEE Xplore.

Liu, Xu, Zhu, Hanhao, Song, Weihua, Wang, Jiahui, Yan, Lengleng, and Wang, Kelin (2024). Improved VGG-16 Model for Acoustic Image Recognition in SAR. IEEE Xplore.

Aji, Andi Alfian Kartika, Sebastian, Matthew, Alexander, Francis, Tsai, Po-Ting, and Lee, Min-Fan Ricky (2024). Neural Network-Assisted Observer for ASV Tracking. IEEE Xplore.

Soorki, Mehdi Naderi, Aghajari, Hossein, Ahmadinabi, Sajad, Babadegani, Hamed Bakhtiari, Chaccour, Christina, and Saad, Walid (2024). LoRa-based SAR Using Meta-RL for UAVs. IEEE Xplore.

Kozlov, Mykyta S., and Malakhov, Eugene V. (2024). Complex Multi-level Drone System for Human Recognition in SAR. IEEE Xplore.

Dousai, Nayee Muddin Khan, and Loncaric, Sven (2022). Detecting Humans in SAR Based on Ensemble Learning. IEEE Xplore.