DOI: 10.2507/36th.daaam.proceedings.xxx

APPLICATION POTENTIAL OF YOLO-BASED REAL-TIME HUMAN DETECTION SYSTEMS IN SEARCH AND RESCUE OPERATIONS

Danijel Zelenika, Željko Marušić, Markić Ivan

This Publication has to be referred as: Zelenika, D[anijel]; Marušić, Ž[eljko] & Markić, I[van] (2025). Title of Paper, Proceedings of the 36th DAAAM International Symposium, pp.xxxx-xxxx, B. Katalinic (Ed.), Published by DAAAM International, ISBN 978-3-902734-xx-x, ISSN 1726-9679, Vienna, Austria
DOI: 10.2507/36th.daaam.proceedings.xxx

Abstract

In Search and Rescue (SAR) operations, rapid and accurate human detection in remote or hazardous areas is critical for saving lives. This study explores the application of modern computer vision algorithms that enable automated, real-time human detection. The focus is placed on deep learning methods, particularly convolutional neural networks (CNNs), with an emphasis on models such as YOLO (You Only Look Once). Although other models may offer higher accuracy, their computational demands pose challenges for real-time processing. The proposed approach was tested on aerial imagery from well-known HERIDAL dataset. The HERIDAL dataset primarily contains simulated rescue scenarios in mountainous and rural areas. The results demonstrate that the model achieves a recall of 85.6% at an input image resolution of 1024px, despite the input image resolution being reduced by nearly four times. Furthermore, the real-time detection can be achieved without significant loss of accuracy, even under low-visibility conditions and complex terrain. Overall, the findings confirm the potential of such systems for practical use in real-world SAR missions.

Keywords: computer vision, human detection, real-time processing, search and rescue missions, convolutional neural networks.

1. Introduction

In emergency situations such as natural disasters, accidents, or missing person incidents, the ability to rapidly and accurately locate individuals can be the determining factor between life and death. Traditional search methods rely on the deployment of large numbers of rescuers, search dogs, and various ground and aerial vehicles, making such operations logistically demanding, costly, and time-consuming. In response to these challenges, unmanned aerial vehicles (UAVs) have become increasingly utilized in SAR missions due to their capability to swiftly survey extensive areas and capture aerial imagery that supports the localization of victims.

However, a significant challenge remains, as the analysis of footage captured by UAVs still largely depends on human interpretation, which makes the process slow and prone to errors [1]. This creates an opportunity for the integration of artificial intelligence, particularly deep learning techniques, which enable the automation of human detection in complex environments. Notably, models from the YOLO family, known for their high-speed and high-accuracy object detection capabilities [2], have demonstrated substantial potential for application in SAR operations.

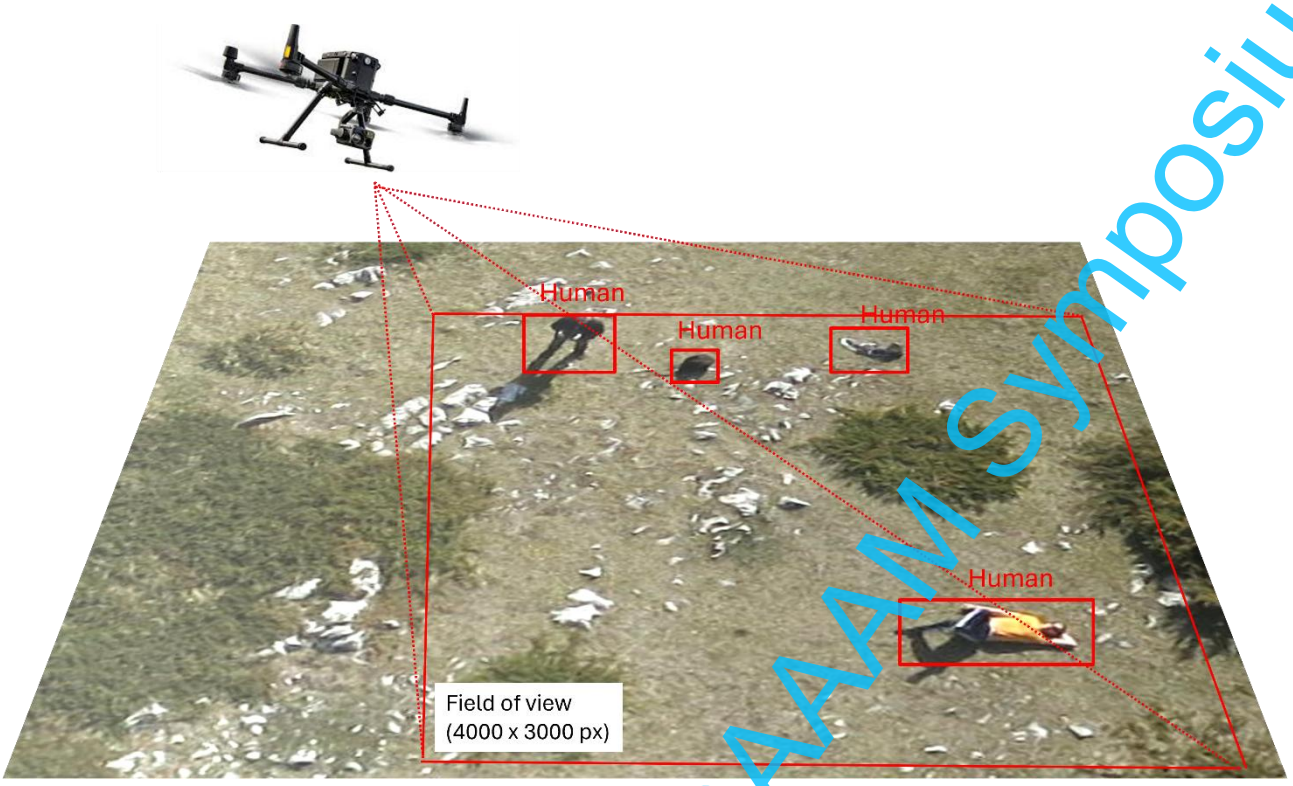

Fig. 1. Examples of human detection on the HERIDAL dataset using the YOLOv11 model.

The objective of this paper is to present the concept of real-time automatic human detection, examine existing methodologies, and discuss the advantages and limitations of such systems in practical SAR scenarios. This paper is structured as follows. Chapter 2 provides an overview of previous research in the field of human detection for SAR operations, presenting related works that address this challenge, the methodologies applied, and the results achieved by various authors. Chapter 3 introduces the YOLOv11 model architecture and describes the datasets used in this study, which serve as the foundation for model training and evaluation. Chapter 4 presents and analyzes the experimental results, focusing on detection accuracy, model performance, and limitations observed during testing. Finally, Chapter 5 summarizes the key findings and provides conclusions and recommendations for future research in the domain of UAV-based human detection systems for SAR applications.

2. Related work

Human detection from aerial imagery represents a highly complex problem in the field of computer vision, primarily due to the significant variability of objects caused by different poses in which a missing or injured person may be found, variations in clothing color, lighting conditions, and background complexity. Traditional object detection approaches have generally been designed for relatively large objects with clearly defined edges and contours. However, in images captured by UAVs, the objects of interest are typically very small, and image instability caused by drone movement further complicates the task.

Conventional object detection methods mainly rely on object size, texture, and shape, whereas modern approaches often employ bimodal systems that combine thermal and optical imagery to improve detection accuracy. Examples include the works of Gaszczak and Breckon [3] and Rudol and Doherty [4], where thermal images are used to identify regions of elevated temperature corresponding to human silhouettes. At the same time, the optical spectrum is analyzed using cascades of boosted classifiers with Haar-like features. Similar approaches have also been applied to the detection of swimmers in open water environments using unsupervised learning techniques [5], whereas studies [6], [7] have proposed models that utilize pyramidal feature extraction in SSD architectures or a combination of color and depth information for human detection.

The use of transfer learning has proven to be an extremely effective strategy in situations where the number of labeled images is limited, which is a common challenge in the analysis of satellite and aerial imagery. Since collecting and manually annotating such data is both time-consuming and resource-intensive, transfer learning enables the reuse of existing pre-trained neural networks as a foundation for further training on domain-specific datasets. This approach significantly improves model accuracy and generalization capabilities without the need to generate large quantities of new annotations or to train models from scratch [8]. For instance, in [9], a model for human detection on aerial images

was first trained on a large, general-purpose dataset encompassing a wide range of visual characteristics, and then fine-tuned on a smaller, domain-specific dataset, achieving high performance and better adaptation to the target environment. Similarly, the study in [10] demonstrates that models utilizing transfer learning on larger and more generic datasets can outperform those trained solely on smaller, specialized datasets. These findings confirm that transferring knowledge from large-scale models enables more efficient learning in resource-constrained scenarios, establishing transfer learning as a key technique in the field of aerial and satellite image analysis.

In Mediterranean regions, the application of thermal cameras during summer months is often constrained by high ambient temperatures, which significantly reduce the temperature contrast between the human body and the surrounding environment. As a result, within the IPSAR project, optical cameras were employed as the primary sensing modality for image acquisition. The image processing workflow incorporated a two-stage mean-shift segmentation algorithm optimized for small image segments, followed by a heuristic analysis stage designed to enhance object distinction and reduce noise. This approach effectively minimizes computational complexity while maintaining a high level of detection accuracy.

Subsequent research efforts expanded upon this foundation by comparing system performance on compressively reconstructed versus original images[11], as well as evaluating the impact of various salient detection algorithms [12] on detection reliability. Additionally, hybrid approaches combining salient detection techniques with convolutional neural network (CNN)-based models [13] have demonstrated promising potential for improving detection robustness and accuracy. However, these methods still face challenges related to false alarm rates and sensitivity under variable environmental conditions, indicating the need for further optimization and adaptive model design in real-world SAR scenarios.

Studies [14] present comparative analyses of the Score Map to ROI algorithm based on deep neural network performance and human expert accuracy across various types of SAR missions, where the Score Map to ROI algorithm have achieved substantially higher accuracy than human evaluators. In the same study, the best results across the four evaluated missions were achieved in experiments where human experts performed detection with the assistance of the Score Map to ROI algorithm. This hybrid approach demonstrated that combining human intuition with algorithmic guidance can significantly enhance detection accuracy and reduce false positives in complex SAR scenarios.

3. Model architecture and Datasets

This chapter focuses on the fundamental components that underpin this research's experimental framework: the YOLOv11 object detection architecture and the aerial HERIDAL image datasets used for model training and evaluation. Collectively, these elements form the technical foundation for developing and accessing automated human detection systems in SAR scenarios.

The YOLOv11 model represents the latest generation of the YOLO family of real-time object detection algorithms. Building on the architecture of YOLOv8, this version integrates several structural and performance improvements aimed at enhancing detection accuracy, computational efficiency, and adaptability across diverse application domains. The model is available in several size variants allowing researchers to balance accuracy and efficiency based on the available computational resources. In this study, the YOLOv11n (nano) model was employed. The model processes an input image of 640×640 pixels in only 8.64 milliseconds on the NVIDIA Jetson Orin NX (16 GB), allowing the system to achieve approximately 116 frames per second (FPS). This exceptional performance underscores the model's suitability for deployment in time-critical applications, such as Search and Rescue (SAR) missions, where rapid detection and low latency are essential for real-time decision-making.

As the foundation for model training and testing, the Heridal [13] dataset was employed. This database consists of a collection of aerial photographs captured by drones, depicting both rural and urban locations across Bosnia and Herzegovina. For the purposes of this study, a subset of images was selected from 11 specific locations: Blidinje, Čapljina, Goranci, Kupres, Ljubuški, Medugorje, Posušje, Rakitno, Široki Brijeg, Stolac, and Velež. The dataset's diversity, including both rural and urban environments, and its high-resolution imagery, make it a valuable resource for developing effective and reliable detection systems for real-world SAR scenarios.

The Heridal dataset contains approximately 1,700 high-resolution images (4000 × 3000 pixels), which allows for detailed analysis of complex scenes. The complete dataset was divided into training and validation sets. To ensure reliability and objective evaluation, the data were randomly split in an 80:20 ratio. This approach ensures that most of the data is used for model learning, while maintaining a representative and unbiased subset for assessing model performance on unseen samples. This dataset has been widely used in research focused on human detection in aerial imagery, particularly in the context of SAR operations. It enables the development and evaluation of various deep learning models, including YOLO-based models for real-time detection, ensemble approaches to improve detection robustness, and transfer learning techniques that leverage pre-trained models adapted to the specific characteristics of aerial images.

4. Results

The proposed model was trained on images with resolutions of 640px and 1024px. The experimental results show that, in both experiments, a stable and prosperous learning process was achieved over 50 epochs, as evidenced by the continuous reduction in the loss function and a significant improvement in performance metrics, both in precision and

recall, as summarized in Table 1. The key metric, mAP@0.5, reached a value of 66.11% at the end of training on 640px images, and 78.33% on 1024px images, indicating a solid foundational ability of the model to identify objects and place bounding boxes with acceptable accuracy.

Despite this success, a more detailed analysis reveals notable weaknesses. The model's performance drops sharply at stricter IoU thresholds, as reflected by the low mAP@0.5-0.95 values of only 26.60% on 640px images and 40.66% on 1024px images. This indicates a significant issue with object localization precision, i.e., placing highly accurate bounding boxes. However, for application in SAR missions, precise bounding box placement is not critical.

Epoch	Precision		Recall		mAP50		mAP50-95	
	640px	1024px	640px	1024px	640px	1024px	640px	1024px
10	51.20	52.65	44.69	54.07	41.64	47.12	13.37	19.91
20	56.01	56.88	51.85	56.60	47.92	62.60	16.16	27.70
30	54.92	53.03	55.18	55.93	49.38	65.57	17.13	29.94
40	57.08	72.69	61.11	70.74	58.46	71.65	19.97	30.41
50	64.53	74.85	68.05	78.89	66.12	78.33	26.60	40.66

Table 1. Performance metrics during training on images with resolutions of 640 and 1024px per epoch.

For model validation, a set of 147 images from the HERIDAL database was selected, containing 270 instances of the "Human" class. Fig. 2 shows an example of successful human detections in a very complex environment, demonstrating the model's ability to accurately identify targets despite challenging background conditions. Two experiments were conducted. In the first experiment, the input image to the model was 640 px, and in the second, 1024 px, given that the original images from the HERIDAL database have dimensions of 4000×3000 pixels. The results of these experiments are presented in Table 2.

Precision		Recall		F1	
640px	1024px	640px	1024px	640px	1024px
71.9	68.8	63.3	85.6	67.3	76.2

Table 2. Comparison of performance metrics for image resolutions of 640 and 1024px.

The analysis of the results on validation set indicates a notably high proportion of false negative detections. Specifically, 99 out of 270 actual instances of the "Human" class were missed in the 640px images. This elevated miss rate, which directly corresponds to a recall of 63.3%, represents the primary limitation of the model. In contrast, the recall on the 1024px images reaches 85.6%, marking a substantial improvement of 22.3%. This significant increase demonstrates that the model benefits considerably from higher-resolution input, leading to more accurate detection of actual objects.

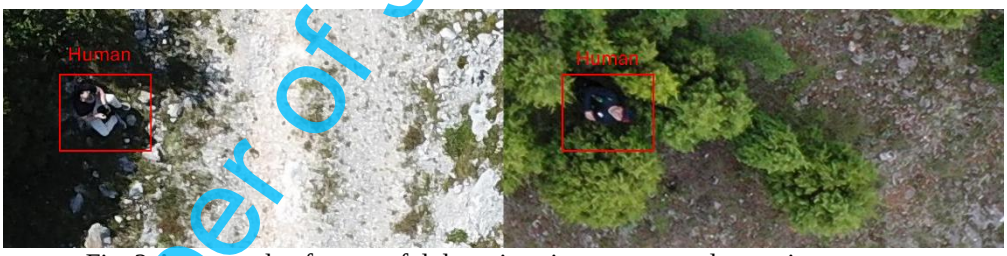

Fig. 2 An example of successful detections in a very complex environment

While the transition to higher-resolution images improved recall, it also increased false positive detections, rising from 67 to 105. However, the overall detection performance, as measured by the F1 score, and the considerable improvement in recall outweighs this drawback. The tendency toward false positives suggests that the model adopts a more "aggressive" detection strategy, aiming to identify every potential object, a behavior commonly observed in models optimized for high recall.

The disparity in false negatives, 39 for 1024px images versus 99 for 640px images, highlights one of the model's key challenges. A detailed examination reveals that the root cause of these limitations lies in the dataset's inherent characteristics. Analysis of the bounding box size distribution confirms that the model is trained predominantly on extremely small objects. Detecting and localizing such small objects is an inherently difficult task, which partially explains the reduced recall at lower resolutions and the increased number of false negatives.

These observations underscore the importance of image resolution in object detection tasks and suggest that higher-resolution inputs enable the model to capture fine-grained features necessary for accurate localization and classification better.

5. Conclusion

This study highlights the critical role of input image resolution in achieving accurate human detection for Search and Rescue (SAR) operations. The results demonstrate that the model's recall improves markedly from 63.3% at 640 px to 85.6% at 1024 px, underscoring the importance of high-resolution inputs for reliable detection. While higher resolutions led to a slight increase in false positive detections, the overall F1 score of 85.6% and substantial gain in recall of 22.3% validate the effectiveness of this approach. Error analysis reveals that the primary challenge lies in detecting very small objects, as represented in the HERIDAL dataset. These limitations are particularly evident at lower resolutions, where a higher incidence of false negatives reduces overall performance.

Despite the increase in false positives, the model adopts a more "aggressive" detection strategy aimed at identifying all relevant objects—a strategy particularly beneficial in SAR applications, where maximizing the detection of actual targets outweighs the cost of occasional false alarms. In summary, these findings confirm that optimizing input resolution and carefully balancing recall and precision can significantly enhance model performance, supporting the practical deployment of human detection systems in complex and hazardous environments.

One of the future goals of this research is to create a more robust image dataset that includes diverse SAR scenarios, such as poor daytime light conditions or snow and other challenging environments, which are currently underrepresented in the official HERIDAL dataset. Another goal is to integrate multiple sensors to improve detection in difficult conditions, such as low-light or poor daytime illumination.

6. References

[1] A. A. Bany Abdelnabi and G. Rabadi, "Human Detection From Unmanned Aerial Vehicles' Images for Search and Rescue Missions: A State-of-the-Art Review," *IEEE Access*, vol. 12, pp. 152009–152035, 2024, doi: 10.1109/ACCESS.2024.3479988.

[2] M. Abdank, M. Aburaia, and W. Woeber, "USING COLOUR-BASED OBJECT DETECTION FOR PICK AND PLACE APPLICATIONS," in *Annals of DAAAM and Proceedings of the International DAAAM Symposium*, DAAAM International Vienna, 2021, pp. 536–541. doi: 10.2507/32nd.daaam.proceedings.077.

[3] A. Gaszczak, T. P. Breckon, and J. Han, "Real-time people and vehicle detection from UAV imagery," in *Intelligent Robots and Computer Vision XXVIII: Algorithms and Techniques*, 2011, p. 78780B. doi: 10.1117/12.876663.

[4] P. Rudol and P. Doherty, "Human body detection and geolocalization for UAV search and rescue missions using color and thermal imagery," in *IEEE Aerospace Conference Proceedings*, 2008. doi: 10.1109/AERO.2008.4526559.

[5] E. Lygouras, N. Santavas, A. Taitzoglou, K. Tarchanidis, A. Mitropoulos, and A. Gasteratos, "Unsupervised Human Detection with an Embedded Vision System on a Fully Autonomous UAV for Search and Rescue Operations," *Sensors*, vol. 19, no. 16, 2019, doi: 10.3390/s19163542.

[6] A. Al-Kaff, M. J. Gómez-Silva, F. M. Moreno, A. de la Escalera, and J. M. Armingol, "An Appearance-Based Tracking Algorithm for Aerial Search and Rescue Purposes," *Sensors*, vol. 19, no. 3, 2019, doi: 10.3390/s19030652.

[7] B. Mishra, D. Garg, P. Narang, and V. Mishra, "Drone-surveillance for search and rescue in natural disaster," *Comput Commun*, vol. 156, pp. 1–10, Apr. 2020, doi: 10.1016/J.COMCOM.2020.03.012.

[8] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," Apr. 2020, [Online]. Available: http://arxiv.org/abs/2004.10934

[9] A. J. Mantau, I. W. Widayat, J.-S. Leu, and M. Köppen, "A Human-Detection Method Based on YOLOv5 and Transfer Learning Using Thermal Image Data from UAV Perspective for Surveillance System," *Drones*, vol. 6, no. 10, 2022, doi: 10.3390/drones6100290.

[10] A. Majidizadeh, H. Hasani, and M. Jafari, "Transfer Learning Framework for Semantic Segmentation of High-Resolution UAV based images in Urban Area," *Journal of Geospatial Information Technology*, vol. 10, no. 4, 2023, doi: 10.61186/jgit.10.4.87.

[11] J. Musić, I. Orović, T. Marasović, V. Papić, and S. Stanković, "Gradient Compressive Sensing for Image Data Reduction in UAV Based Search and Rescue in the Wild," *Math Probl Eng*, vol. 2016, p. 6827414, 2016, doi: 10.1155/2016/6827414.

[12] S. Gotovac, V. Papić, and Ž. Marušic, "Analysis of saliency object detection algorithms for search and rescue operations," in *2016 24th International Conference on Software, Telecommunications and Computer Networks, SoftCOM 2016*, 2016. doi: 10.1109/SOFTCOM.2016.7772118.

[13] D. Božić-Štulić, Ž. Marušić, and S. Gotovac, "Deep Learning Approach in Aerial Imagery for Supporting Land Search and Rescue Missions," *Int J Comput Vis*, 2019, doi: 10.1007/s11263-019-01177-1.

[14] S. Gotovac, D. Zelenika, Ž. Marušić, and D. Božić-Štulić, "Visual-Based Person Detection for Search-and-Rescue with UAS: Humans vs. Machine Learning Algorithm," *Remote Sens (Basel)*, vol. 12, no. 20, 2020, doi: 10.3390/rs12203295.