VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.8002

JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

COMPARATIVE BENCHMARKING OF YOLO11 FOR UAV-BASED HUMAN DETECTION IN SIMULATED DISASTER SCENARIOS

A. Ahmad Fadil[1*]; Jumadi Mabe Parenreng[1]; Muhammad Fardan[1]; Muhammad Fajar B[1]; Ana Sulistiana Alwi[1]

Computer Engineering[1]
Universitas Negeri Makassar, Makassar, Indonesia[1]
https://unm.ac.id[1]
fadilandi57@gmail.com[*], jparenreng@unm.ac.id, muhammadfardhann123@gmail.com, fajarb@unm.ac.id, anasulistiana563@gmail.com

(*) Corresponding Author
(Responsible for the Quality of Paper Content)

Abstract— *Search and Rescue (SAR) operations require rapid and reliable identification of survivors in disaster-affected areas. This study presents an empirical benchmark of YOLO11m for UAV-oriented human detection in simulated disaster scenarios. The model was trained and evaluated using the synthetically generated C2A: Human Detection in Disaster Scenarios dataset and compared with YOLOv8m, YOLOv9m, and YOLOv10m under consistent training configurations. The experiments were conducted at a resolution of 640 × 640 pixels using mixed-precision training on dual NVIDIA T4 GPUs. YOLO11m achieved an mAP@50 of 0.850, an mAP@50–95 of 0.612, a Precision of 0.880, a Recall of 0.799, and a peak F1-score of 0.830 at a confidence threshold of 0.376. The model also recorded an average inference latency of 5.5 ms per image in the test environment. Compared with the evaluated medium-scale YOLO variants, YOLO11m achieved the highest mAP@50–95 while maintaining moderate parameter and computational requirements. These results indicate that YOLO11m provides a promising accuracy–efficiency baseline for UAV-based SAR research. However, because the evaluation relies on synthetic data and does not include onboard inference or physical UAV field trials, further validation using real-world aerial disaster imagery is required before operational deployment.*

Keywords: *Computer Vision, Human Detection , Machine Learning, UAV, YOLO*

Intisari— *Operasi Pencarian dan Penyelamatan (Search and Rescue atau SAR) memerlukan identifikasi korban selamat yang cepat dan andal di wilayah terdampak bencana. Penelitian ini menyajikan tolok ukur empiris YOLO11m untuk deteksi manusia berorientasi UAV dalam skenario bencana tersimulasi. Model dilatih dan dievaluasi menggunakan dataset sintetis C2A: Human Detection in Disaster Scenarios serta dibandingkan dengan YOLOv8m, YOLOv9m, dan YOLOv10m menggunakan konfigurasi pelatihan yang konsisten. Eksperimen dilakukan pada resolusi 640 × 640 piksel dengan pelatihan presisi campuran (mixed-precision training) menggunakan dua GPU NVIDIA T4. YOLO11m mencapai mAP@50 sebesar 0.850, mAP@50–95 sebesar 0.612, presisi sebesar 0.880, recall sebesar 0.799, dan skor F1 tertinggi sebesar 0.830 pada ambang batas kepercayaan (confidence threshold) 0,376. Model ini juga mencatat rata-rata latensi inferensi sebesar 5.5 ms per citra pada lingkungan pengujian. Dibandingkan dengan varian YOLO berskala menengah yang dievaluasi, YOLO11m memperoleh nilai mAP@50–95 tertinggi dengan kebutuhan parameter dan komputasi yang moderat. Hasil ini menunjukkan bahwa YOLO11m dapat menjadi tolok ukur awal yang menjanjikan dalam menyeimbangkan akurasi dan efisiensi untuk penelitian SAR berbasis UAV. Namun, karena evaluasi masih menggunakan data sintetis serta belum mencakup inferensi pada perangkat UAV dan pengujian lapangan menggunakan UAV fisik, validasi lebih lanjut menggunakan citra udara dari kondisi bencana nyata diperlukan sebelum model dapat diterapkan secara operasional.*

Kata Kunci: *Machine Learning, Computer Vision, UAV, YOLO, Human Detection.*

VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.8002

INTRODUCTION

Due to climate change and associated environmental factors, the frequency of natural disasters such as floods, earthquakes, and hurricanes has increased, causing significant harm to human life and urban infrastructure [1], [2]. These catastrophic events disrupt the essential economic and social functioning of metropolitan areas [3]. Consequently, Search and Rescue (SAR) operations are of utmost importance, yet they often face operational bottlenecks that delay response times. Successful SAR missions require an immediate and rapid response to the needs of afflicted communities [4].

Unmanned Aerial Vehicles (UAVs), commonly known as drones, are increasingly deployed to mitigate these challenges. While initially developed for military purposes, significant research has focused on leveraging their autonomous capabilities for civilian applications [5]. In SAR contexts, drones can quickly survey large, complex areas, enabling the early location of individuals trapped by disasters [6].

To maximize the utility of drones, machine learning algorithms are embedded to facilitate autonomous object detection. The "You Only Look Once" (YOLO) architecture is particularly prominent because it treats object detection as a single regression problem [7]. Unlike traditional models that utilize multi-stage processes, such as extracting edge information over multiple frames, YOLO processes full images through a single neural network. This efficiency allows it to process streaming video signals in real time, making it superior to traditional methods for UAV-based SAR missions [8], [9], [10], [11].

However, using UAVs for human detection remains challenging due to small object sizes, fluctuating light conditions, and complex backgrounds [12]. Valarmathi et al. (2023) addressed this using YOLOv3, achieving 94.6% accuracy in human activity recognition [13]. Furthermore, Mantau et al. (2022) integrated RGB and thermal images using YOLOv5, achieving Mean Average Precision (mAP) scores of 91.5% and 93.2% respectively, highlighting the benefit of multimodal sensors [14].

Recent literature continues to benchmark YOLO's evolution. Dumenčić et al. (2025) achieved a 73% mAP using a UAV-adapted YOLOv8, demonstrating high locational reliability [15]. Bachir and Memon (2024) developed a custom YOLOv5 system reaching 96.9% mAP, outperforming Faster R-CNN [16]. Similarly, Wu et al. (2024) introduced the YOLOv5_mamba, which integrates YOLOv8 features to improve small object detection on the VisDrone dataset [17].

Despite the progress of YOLO-based object detection for UAV applications, the performance of YOLO11m for human detection in simulated disaster scenarios has not been sufficiently characterized through a controlled comparison with earlier medium-scale YOLO variants. Previous studies have evaluated different YOLO generations using varying datasets, model scales, sensor modalities, and experimental settings, which limits direct interpretation of whether performance improvements are attributable to the model architecture or to differences in evaluation conditions. To address this gap, this study benchmarks YOLO11m against YOLOv8m,

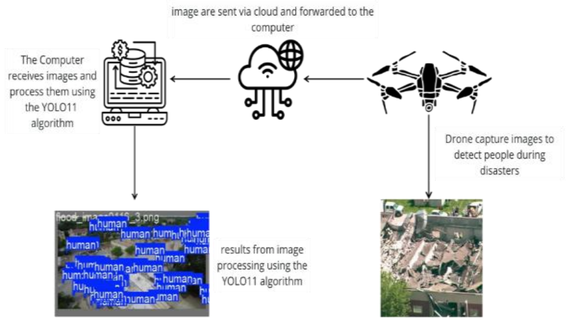
YOLOv9m, and YOLOv10m using the synthetically generated C2A: Human Detection in Disaster Scenarios dataset. All models are initialized with pretrained weights and trained under consistent experimental configurations, including the same dataset split, image resolution, batch size, number of epochs, and evaluation metrics. Rather than proposing a new architecture, dataset, or operational UAV deployment system,

This study provides a controlled empirical assessment of the accuracy-efficiency trade-off of YOLO11m as a baseline for future UAV-based SAR research. Because the experiments rely on synthetic data and do not include onboard inference or physical UAV field trials, further validation using real-world aerial disaster imagery is required before operational deployment.

MATERIALS AND METHODS

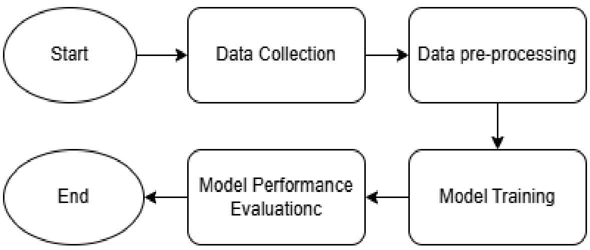
This study employed a quantitative experimental approach to evaluate the performance of the YOLO11m algorithm for human detection in UAV-oriented disaster scenarios. The experiments were conducted using secondary data from the *C2A: Human Detection in Disaster Scenarios* dataset, which contains annotated images representing various disaster conditions, including floods and landslides.

The methodological workflow consists of four main stages: data collection, data preprocessing, model training, and model performance evaluation. The UAV-based image acquisition process illustrated in Figure 1 represents a conceptual deployment scenario for future field implementation rather than an operational component of the present experiment.



Source: (Research Results, 2025)
Figure 1. Conceptual Workflow Of UAV-Based Human Detection in Disaster Scenarios

Figure 1 illustrates a conceptual workflow for the future deployment of UAV-based human detection in disaster scenarios. In this conceptual setting, a UAV equipped with a camera captures aerial images of disaster-affected areas. The collected images are transferred to a processing environment and analyzed using the YOLO11 algorithm to detect human objects. However, the present study does not implement or evaluate image acquisition using a physical UAV, cloud-based data transmission, or onboard inference. Instead, the experiments are conducted using secondary data from the synthetically generated C2A: Human Detection in Disaster Scenarios dataset, which contains annotated images representing various disaster conditions, including floods, fires, and earthquakes. The actual experimental procedure consists of four main stages: secondary data collection, data preprocessing, model training, and model performance evaluation.



Source: (Research Results, 2025)
Figure 2. Experimental Workflow for Training and Evaluating YOLO11m Using the Secondary C2A Dataset

1. Data Collection

Secondary data from the C2A (Combination to Application) Dataset: Human Detection in Disaster Scenarios is employed in this paper, and it is available on Kaggle (2024). This dataset provides a large number of annotated samples of training deep learning models, it is important to acknowledge that the dataset is synthetically generated through fusion of disaster backgrounds from AIDER [18] and human poses from LSP/MPII-MPHB. The dataset contain 10,215 high-resolution images with over 360,000 labels across five body positions and four disaster types [19]. However, due to its synthetically generated nature, it may not fully represent the visual complexity and variability encountered in real-world UAV disaster imagery, such as unpredictable lighting conditions, heavy occlusions, severe motion blur, or sensor noise. Nevertheless, the C2A dataset remains highly valuable for initial model development and benchmarking due to its scale and diversity. Consequently, any claims regarding operational readiness in this study are strictly treated as theoretical baselines that necessitate subsequent physical field validation.

2. Data Pre-processing

Given the C2A Dataset's robust structure and detailed annotations, preprocessing was minimal. The data was partitioned into balanced training, validation, and test subsets using a ratio of 60%, 20%, and 20%, respectively. Preparation for YOLO11 involved standardizing image dimensions to 640×640 pixels, converting formats, and normalizing pixel values to facilitate convergence. Consistency checks confirmed valid image-label pairings, preventing data errors. Consequently, the dataset was directly suitable for training without the need for additional annotation.

3. Model Training

The YOLO11m model was selected for its superior efficiency, offering higher mAP scores with approximately 22% fewer parameters than YOLOv8m. To ensure rigorous methodological reproducibility, the proposed YOLO11m model and all comparative baseline models (YOLOv8m, YOLOv9m, YOLOv10m) were initialized with pretrained weights and trained under identical configurations. Training utilized the C2A Dataset over 50 epochs at a resolution of 640×640 pixels with a fixed batch size of 16. The process was executed on the Kaggle platform using dual NVIDIA T4 GPUs with Automatic Mixed Precision (AMP) enabled to accelerate convergence. To guarantee deterministic and reproducible results, the random seed was fixed at 0. The Ultralytics training environment was configured with an initial learning rate of 0.01, a momentum of 0.937, a weight decay of 0.0005, and an IoU threshold of 0.7. Furthermore, the model utilized a robust spatial augmentation strategy consisting of mosaic augmentation (1.0 probability) that was specifically disabled during the final 10 epochs, alongside horizontal flipping

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.8002

(0.5), spatial scaling (0.5), translation (0.1), and random erasing (0.4). Validation metrics including loss, Precision, Recall, and mAP were continuously monitored to prevent overfitting, and the checkpoint with the highest validation mAP was selected as the final model for testing.

4. Model Evaluation

The model evaluation assessed the YOLO11m algorithm using the C2A test set. Performance was measured via Precision, Recall, and mean Average Precision (mAP). Equation (1) defines Precision as the ratio of correct detections (True Positives) to total positive predictions [20].

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

Equation (2) defines Recall, which quantifies the proportion of actual human objects successfully identified [21].

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

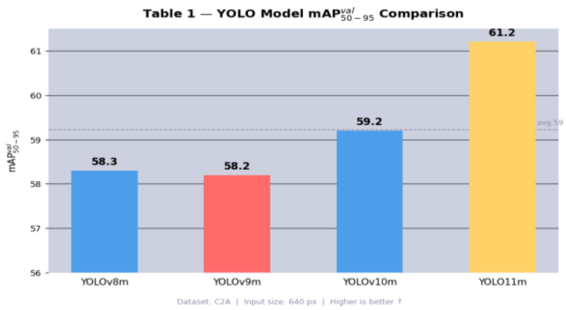
Equation (3) defines mean Average Precision (mAP) as the average of the Average Precision (AP) values across all evaluated classes. Each AP value is calculated from the area under the Precision-Recall curve [22].

$$mAP = \frac{1}{n}\sum_{i=1}^{i=n} AP_i \quad (3)$$

RESULTS AND DISCUSSION

Comparative analysis was conducted between YOLOv8m, YOLOv9m, YOLOv10m, and YOLO11m using the C2A dataset to evaluate the trade-off between detection accuracy and computational efficiency. The assessment focused on mean Average Precision (mAP@0.5:0.95) relative to inference time. As illustrated in Figure 3, the proposed YOLO11m model demonstrates a superior balance of performance compared to its predecessors.

Figure 3 shows that YOLO11m provides the most favorable accuracy-efficiency trade-off among the evaluated models. It achieves the highest mAP@0.5:0.95 while maintaining moderate computational requirements. YOLOv10m requires fewer FLOPs but produces slightly lower detection accuracy, indicating that a reduction in computational complexity may involve a minor performance trade-off. In contrast, YOLOv8m and YOLOv9m require greater computational resources without surpassing the accuracy achieved by YOLO11m.



Source: (Research Results, 2025)

Figure 3 Comparison mAP@0.5:0.95 Values Among Medium-Scale YOLO Variants on the C2A Dataset.

These findings suggest that YOLO11m is a suitable empirical baseline for further UAV-based human detection research. However, additional experiments using repeated training runs and real-world aerial imagery are required to confirm the robustness and practical relevance of this comparison. Further quantitative details regarding model parameters and FLOPs are presented in Table 1.

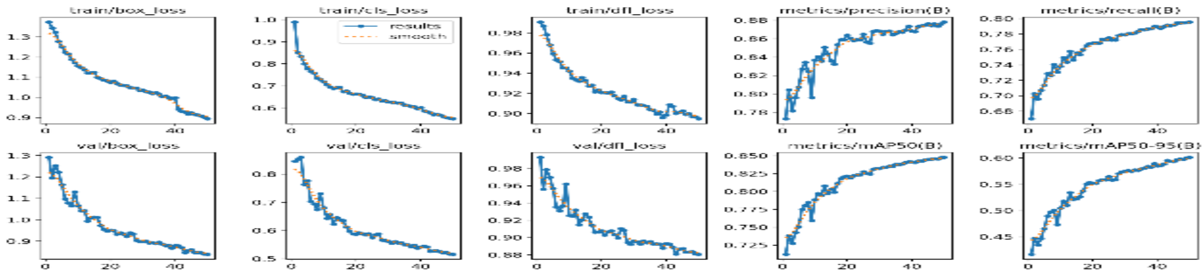
Table 1 Performance Comparison of the YOLOm Using C2A Dataset Testing

Model	Size	mAP^{val}_{50-95}	Param(M)	Latency (ms)	Flops
YOLOv8m	640	0.583	25.8	11.4	78.7 G
YOLOv9m	640	0.582	20.1	5.1	76.5 G
YOLOv10m	640	0.592	15.3	4.8	58.9 G
YOLO11m	640	0.612	20.3	5.5	67.6 G

Source: (Research Results, 2025)

Table 1 shows that YOLO11m achieves the highest detection accuracy among the evaluated medium-scale YOLO variants, with an mAP@0.5:0.95 of 0.612. The model requires 20.3 million parameters and 67.6 GFLOPs, while recording an average inference latency of 5.5 ms per image at a resolution of 640 × 640 pixels. Although YOLOv10m achieves a lower latency of 4.8 ms and requires fewer computational resources, its mAP@0.5:0.95 is slightly lower at 0.592. Meanwhile, YOLOv8m and YOLOv9m do not surpass the accuracy of YOLO11m. These results indicate that YOLO11m provides a competitive balance between detection accuracy and computational efficiency under the specified test conditions. However, further onboard inference testing and physical UAV field trials are required to evaluate its practical suitability for time-critical SAR missions.

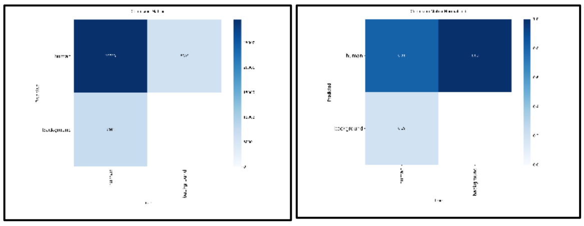
VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.8002



Source: (Research Results, 2025)

Figure 4 A Visual of the Training and Validation Performance for the Best Configuration YOLO11m

Figure 4 illustrates the YOLO11m training trajectory over 50 epochs, showing stable convergence without overfitting. Training losses (box, classification, DFL) decreased consistently to 0.86, 0.59, and 0.91, respectively. Validation losses mirrored this trend, dropping to 0.78, 0.53, and 0.90, indicating strong generalization. Detection metrics improved significantly: Precision rose to 0.880, Recall to 0.799, mAP@50 to 0.85, and mAP@50-95 to 0.612. The minimal gap (<0.05) between training and validation losses confirms the model's robustness and consistency.

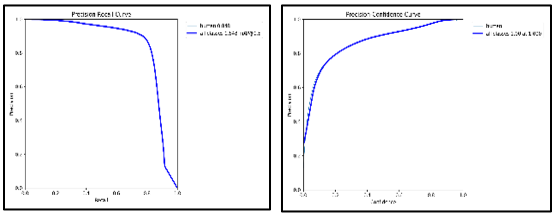


(a) (b)

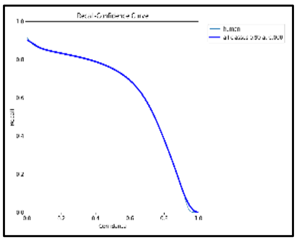
Source: (Research Results, 2025)

Figure 5 Absolute (a) and normalized (b) confusion matrices of YOLO11m on the C2A Dataset at a confidence threshold of 0.25

Figure 5 presents the absolute and normalized confusion matrices for YOLO11m, illustrating its classification performance at a fixed confidence threshold. The clustering of predictions along the diagonal indicates a high-performing classifier. Quantitatively, the model recorded 29,733 true positives (TP) for the 'human' class, alongside 5,899 false positives (FP) and 7,045 false negatives (FN). The normalized confusion matrix reinforces this interpretation, demonstrating a true positive detection rate (recall) of 0.81 for the 'human' class. Furthermore, the matrix displays a normalized value of 1.00 in the top-right quadrant; because the matrix is column-normalized by true labels, this indicates that 100% of the background false positives were assigned to the singular 'human' class, which is an expected mathematical outcome for single-class detection models rather than a metric of precision. When evaluated across all confidence thresholds to maximize the global F1-score, the aggregated model achieves a consistent overall Precision of 0.878. This confirms YOLO11m's strong capability to differentiate between human subjects and the environment. Such discriminatory power is critical for UAV-based SAR missions; by minimizing false alarms, the system conserves operational resources and ensures rescue teams remain focused on genuine threats.



(a) (b)



(c)

Source: (Research Results, 2025)

Figure 6 Performance curves of YOLO11m on the C2A Dataset: (a) Precision–Recall, (b) Precision–Confidence, and (c) Recall–Confidence The Precision-Recall curve reports detection performance at an IoU threshold of 0.50.

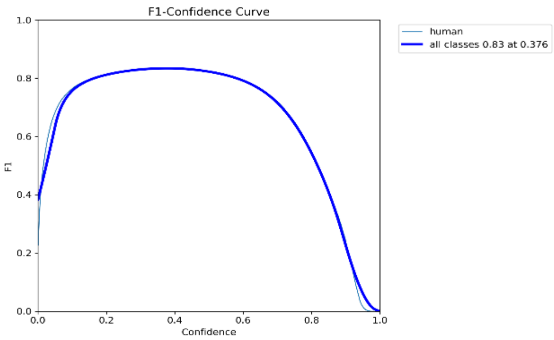
Figure 6 illustrates the performance curves of YOLO11m on the C2A Dataset, specifically the Precision-Recall, Precision-Confidence, and Recall-Confidence variations. The Precision-Recall curve yields a mean Average Precision (mAP@0.5) of 0.85, highlighting strong overall detection capabilities. Furthermore, the Recall-Confidence curve demonstrates that recall peaks at 0.90 at near-zero thresholds, indicating the model detects

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.8002

approximately 90% of individuals when sensitivity is maximized.

As the confidence threshold rises, recall gradually declines while precision increases, reflecting the standard trade-off where higher selectivity reduces false positives but increases the risk of missed detections. In simulated Search and Rescue (SAR) contexts, where missing a victim has severe consequences, high recall is prioritized. The dataset analysis suggests that a confidence threshold between 0.3 and 0.5 provides an optimal mathematical balance, maintaining recall above 0.75 while effectively controlling false alarms. While these findings imply that operators could hypothetically lower thresholds in cluttered or low-visibility environments to prioritize detection, these adjustments are based solely on static dataset evaluations. The practical efficacy of such threshold adaptability for dynamic UAV-based disaster response remains a theoretical projection requiring future mission-level field validation.



Source: (Research Results, 2025)
Figure 7 F1-Confidence curve of YOLO11m on the C2A test set. The highest F1-score of 0.83 is achieved at a confidence threshold of 0.376

Figure 7 depicts the F1-score variation against confidence thresholds for the YOLO11m model trained on the C2A dataset. The F1-score, representing the harmonic mean of precision and recall, balances the trade-off between false positives and false negatives. As the confidence threshold rises, the model applies stricter criteria, increasing precision at the expense of recall. The resulting curve follows a bell-shaped trajectory, peaking at an optimal confidence threshold of 0.376 with a maximum F1-score of 0.83.

This specific point represents the mathematical optimum for maximizing detection reliability on the C2A dataset, minimizing both false alarms and missed targets. Beyond this peak, the F1-score declines as the loss in recall outweighs gains in precision. Consequently, a threshold near 0.376 serves as a strong baseline for Search and Rescue (SAR) applications. While this suggests the parameter could potentially be adapted to accommodate varying environmental conditions such as lighting or clutter, such operational agility remains speculative at this stage. Ensuring consistent real-time performance in high-stakes disaster scenarios will require extensive field testing to validate how these threshold adjustments behave under physical operational constraints.

(a) (b)

Source: (Research Results, 2025)
Figure 8 Qualitative detection results of YOLO11m on selected C2A test images: (a) ground-truth annotations and (b) predicted bounding boxes generated using a confidence threshold of 0.25.

Figure 8 shows that the predicted bounding boxes align closely with actual labels, confirming accurate localization. Confidence scores ranging from 0.3 to 0.9 indicate detection stability across varying poses, object sizes, and cluttered environments. However, specific challenges persist. In densely populated scenes, the model occasionally produces redundant detections due to overlapping instances; this can be mitigated by adjusting Non-Maximum Suppression (NMS) parameters or implementing Soft-NMS [23], as illustrated in Additionally, performance degrades in low-contrast or heavily shadowed areas. Incorporating data augmentation techniques related to brightness and color variance during training could further enhance robustness under these conditions [24], [25].

Quantitatively, YOLO11m demonstrates superior stability. At the optimal confidence threshold of 0.376, the model achieves a Precision of 0.880, Recall of 0.799, and a peak F1-score of 0.83. A mAP@50 of 0.85 highlights accurate detection capabilities, while a mAP@50–95 of 0.612 reflects robustness under strict Intersection over Union (IoU) criteria. These metrics confirm that YOLO11m effectively balances accuracy and computational speed within the tested parameters, validating its strong theoretical suitability as a foundational algorithm for UAV search and rescue missions, pending verification in complex, real-world disaster environments.

CONCLUSION

Utilizing the "C2A: Human Detection in Disaster Scenarios" dataset, the YOLO11m model demonstrated successful convergence and high operational stability. The evaluation yielded a Precision of 0.880, Recall of 0.799, and an F1-score of 0.83. With a mean Average Precision (mAP@50) of 0.85 and mAP@50–95 of 0.612, the model exhibits robust reliability across diverse disaster environments, including floods, fires, and structural collapses. Furthermore, achieving an average processing speed of 5.5 ms during testing suggests the system possesses the computational efficiency required to theoretically support real-time inference on standard UAV platforms. While minor challenges persist, specifically regarding bright lighting and duplicate detections in crowded scenarios, these can be mitigated through targeted optimization in future work. Ultimately, rather than introducing architectural innovations, this study validates the existing model's balance of accuracy and efficiency on synthetic data, establishing a verified empirical foundation to guide future physical drone deployments and real-world search and rescue field trials.

REFERENCE

[1] Q. An and C. Ochiai, "Urban morphology and disaster risk reduction: A systematic literature review," *Progress in Disaster Science*, vol. 27, p. 100455, Oct. 2025, doi: 10.1016/J.PDISAS.2025.100455.
[2] D. N. Nguyen, Y. Usuda, and F. Imamura, "Gaps in and Opportunities for Disaster Risk Reduction in Urban Areas Through International Standardization of Smart Community Infrastructure," *Sustainability*, vol. 16, no. 21, p. 9586, Nov. 2024, doi: 10.3390/su16219586.
[3] A. Reimuth, M. Hagenlocher, L. E. Yang, A. Katzschner, M. Harb, and M. Garschagen, "Urban growth modeling for the assessment of future climate and disaster risks: approaches, gaps and needs," *Environmental Research Letters*, vol. 19, no. 1, p. 013002, Jan. 2024, doi: 10.1088/1748-9326/ad1082.
[4] T. Khan Mohd, V. Nguyen, T. Hoang, P. M. Zeyede, and B. Abdisa, "Performance and Efficiency Assessment of Drone in Search and Rescue Operation," in *Artificial Intelligence and Machine Learning*, Academy and Industry Research Collaboration Center (AIRCC), Jul. 2022, pp. 1–16. doi: 10.5121/csit.2022.121201.
[5] M. Mbarak, M. H. Alam, and M. Awad, "Unmanned Aerial Vehicle Technologies, Applications, and Regulatory Frameworks: A Scoping Review," *Drones*, vol. 10, no. 5, p. 365, May 2026, doi: 10.3390/drones10050365.
[6] F. Ciccone and A. Ceruti, "Real-Time Search and Rescue with Drones: A Deep Learning Approach for Small-Object Detection Based on YOLO," *Drones*, vol. 9, no. 8, p. 514, Jul. 2025, doi: 10.3390/drones9080514.
[7] T. Diwan, G. Anirudh, and J. V. Tembhurne, "Object detection using YOLO: challenges, architectural successors, datasets and applications," *Multimed. Tools Appl.*, vol. 82, no. 6, pp. 9243–9275, Mar. 2023, doi: 10.1007/s11042-022-13644-y.
[8] M. L. Ali and Z. Zhang, "The YOLO Framework: A Comprehensive Review of Evolution, Applications, and Benchmarks in Object Detection," *Computers*, vol. 13, no. 12, p. 336, Dec. 2024, doi: 10.3390/computers13120336.
[9] W. A. Ezat, M. M. Dessouky, and N. A. Ismail, "Evaluation of Deep Learning YOLOv3 Algorithm for Object Detection and Classification," *Menoufia Journal of Electronic Engineering Research*, vol. 30, no. 1, pp. 52–57, Jan. 2021, doi: 10.21608/mjeer.2021.146237.
[10] G. Lavanya and S. D. Pande, "Enhancing Real-time Object Detection with YOLO Algorithm," *EAI Endorsed Transactions on Internet of Things*, vol. 10, Dec. 2023, doi: 10.4108/eetiot.4541.
[11] M. Nikouei *et al.*, "Small object detection: A comprehensive survey on challenges, techniques and real-world applications," *Intelligent Systems with Applications*, vol. 27, p. 200561, Sep. 2025, doi: 10.1016/j.iswa.2025.200561.
[12] C. Chen *et al.*, "YOLO-Based UAV Technology: A Review of the Research and Its Applications," *Drones*, vol. 7, no. 3, p. 190, Mar. 2023, doi: 10.3390/drones7030190.
[13] B. Valarmathi *et al.*, "Human Detection and Action Recognition for Search and Rescue in Disasters Using YOLOv3 Algorithm," *Journal of Electrical and Computer Engineering*, vol. 2023, pp. 1–19, Mar. 2023, doi: 10.1155/2023/5419384.
[14] A. J. Mantau, I. W. Widayat, J.-S. Leu, and M. Köppen, "A Human-Detection Method Based on YOLOv5 and Transfer Learning Using

VOL. 12. NO. 1 AUGUST 2026
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v12i1.8002

Thermal Image Data from UAV Perspective for Surveillance System," *Drones*, vol. 6, no. 10, p. 290, Oct. 2022, doi: 10.3390/drones6100290.

[15] S. Dumenčić, L. Lanča, K. Jakac, and S. Ivić, "Experimental Validation of UAV Search and Detection System in Real Wilderness Environment," *Drones*, vol. 9, no. 7, p. 473, Jul. 2025, doi: 10.3390/drones9070473.

[16] N. Bachir and Q. A. Memon, "Benchmarking YOLOv5 models for improved human detection in search and rescue missions," *Journal of Electronic Science and Technology*, vol. 22, no. 1, p. 100243, Mar. 2024, doi: 10.1016/j.jnlest.2024.100243.

[17] S. Wu, X. Lu, and C. Guo, "YOLOv5_mamba: unmanned aerial vehicle object detection based on bidirectional dense feedback network and adaptive gate feature fusion," *Sci. Rep.*, vol. 14, no. 1, p. 22396, Sep. 2024, doi: 10.1038/s41598-024-73241-x.

[18] R. A. Nihal, B. Yen, K. Itoyama, and K. Nakadai, "UAV-Enhanced Combination to Application: Comprehensive Analysis and Benchmarking of a Human Detection Dataset for Disaster Scenarios," 2025, pp. 145–162. doi: 10.1007/978-3-031-78341-8_10.

[19] R. A. Nihal, B. Yen, K. Itoyama, and K. Nakadai, "UAV-Enhanced Combination to Application: Comprehensive Analysis and Benchmarking of a Human Detection Dataset for Disaster Scenarios," Aug. 2024, Accessed: Nov. 05, 2025. [Online]. Available: https://arxiv.org/abs/2408.04922

[20] W. Wang, L. Yang, and N. Noguchi, "Development of a grape-harvesting robot using a multi-step detection method based on AI and a position-estimation algorithm," *Smart Agricultural Technology*, vol. 9, p. 100574, Dec. 2024, doi: 10.1016/j.atech.2024.100574.

[21] E. I. Capaldi, "A low-cost wireless extension for object detection and data logging for educational robotics using the ESP-NOW protocol," *PeerJ Comput. Sci.*, vol. 10, p. e1826, Feb. 2024, doi: 10.7717/peerj-cs.1826.

[22] Z. Zhao, L. Alzubaidi, J. Zhang, Y. Duan, and Y. Gu, "A comparison review of transfer learning and self-supervised learning: Definitions, applications, advantages and limitations," *Expert Syst. Appl.*, vol. 242, p. 122807, May 2024, doi: 10.1016/j.eswa.2023.122807.

[23] N. Bodla, B. Singh, R. Chellappa, and L. S. Davis, "Soft-NMS -- Improving Object Detection With One Line of Code," Aug. 2017, Accessed: Nov. 05, 2025. [Online]. Available: https://arxiv.org/abs/1704.04503

[24] G. Li, L. Wen, Z. Huang, R. Xia, and Y. Pang, "Data augmentation and shadow image classification for shadow detection," *IET Image Process.*, vol. 16, no. 3, pp. 717–728, Feb. 2022, doi: 10.1049/ipr2.12377.

[25] Y.-F. Lu, J.-W. Gao, Q. Yu, Y. Li, Y.-S. Lv, and H. Qiao, "A Cross-Scale and Illumination Invariance-Based Model for Robust Object Detection in Traffic Surveillance Scenarios," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 7, pp. 6989–6999, Jul. 2023, doi: 10.1109/TITS.2023.3264573.