Data Science and Engineering (DSE) Record, Volume 6, issue 1.

DEVELOPMENT OF HUMAN DETECTION AND LOCALIZATION FROM IMAGES BY DRONE IN SEARCH OPERATION

Kridsanaphon Suksan [1] and Paskorn Champrasert [2]

[1] Master's Degree Program in Data Science, Chiang Mai University, Chiang Mai, Thailand
[2] Department of Computer Engineering, Faculty of Engineering, Chiang Mai University, Chiang Mai, Thailand
kridsanaphon_suksan@cmu.ac.th

Abstract. This independent study presents a system for object detection and localization using aerial imagery captured by drones in search and rescue operations. Generally, higher drone altitude gives greater area coverage, but reduces detection accuracy. While a lower altitude improves accuracy, but requires more search time. Lacking guidance on optimal altitude information, this study explores the various detection performances at different flight altitudes to enhance operational efficiency. Since altitude impacts both image quality and detection accuracy, image resolution is also examined as a key factor in system performance. The study evaluates the YOLOv11 algorithm for detection in aerial images, using clothing as a human proxy to address ethical and data collection constraints. Performance was assessed using Mean Average Precision, Precision, Recall, and Time along with, derived metrics like Efficiency Score and Missing Rate. The geolocation deviation is also measured. Findings indicated that increasing altitude reduces model performance but can be compensated by using a higher resolution image. For missions requiring high detection accuracy, the lowest altitude flights yield the best results. In contrast, more time-constrained operations can benefit from higher altitude but need more computation resources. In general, the study suggests a flight altitude of 40 meters with 1080×720 resolution as the most efficient altitude. At 40 meters, detection accuracy slightly decreases, but area coverage and computation speed improve significantly by roughly three times with the top Efficiency Score and lowest Missing Rate.

Keywords: Drone, Object detection, Search and Rescue.

1 Introduction

In recent years, the frequency and severity of both natural and man-made disasters have intensified. Robotics was invented to perform tasks that mimic human activity, and by the 1990s, rescue robotics began to assist in Search and Rescue (SAR) operations [1], followed by the development of aerial robotics or Unmanned Aerial Vehicles (UAVs) [2]. Today, UAVs (drones) are recognized as valuable tools for SAR due to their ability to access hard-to-reach areas, navigate complex environments, and provide aerial imagery beyond human perspective.

Drones offer rapid visual coverage in inaccessible or obstructed areas, enhancing SAR missions where timely and accurate victim localization is crucial [3][4]. However, detection in drone imagery poses technical challenges including occlusion, scale variability due to altitude, and false positives caused by similar-looking objects like rocks or shadows. These factors emphasize the need for robust, adaptive detection systems capable of operating under real-world constraints [5].

To support testing while avoiding ethical concerns, this study simulates human detection using clothing. This approach ensures safety, repeatability, and diverse training data collection. Although it lacks human body structure, clothing allows for scalable dataset creation under varying conditions, aiding model generalization.

Despite advances in drone technology, SAR still faces trade-offs between detection accuracy, speed, and location precision. Higher altitudes enable faster area coverage but reduce image detail and GPS precision. Lower flights improve accuracy but slow the process. Current detection systems struggle to balance these constraints effectively. This study addresses this gap by implementing an object detection framework that evaluates how flight altitude affects detection accuracy, speed, and geolocation precision.

The proposed framework utilizes the YOLO (You Only Look Once) object detection system [6], known for its speed and accuracy. Unlike traditional multi-stage models [7], YOLO processes entire images in one pass, making it suitable for real-time applications on compact drone hardware. Its adaptability through pretrained models enhances field readiness without complex setups.

This study aims to develop a practical SAR framework using aerial imagery to detect and geolocate objects via GPS. It examines how flight altitude and image resolution impact detection performance. Since higher altitudes reduce pixel density, larger image resolutions than YOLO's default 640×640 may be needed for accurate results. The goal is to optimize detection accuracy, speed, and operational suitability for real-time SAR applications.

2 2. Literature Review

2.1 Drone Technology in Search and Rescue Operations

Drone technology has significantly enhanced Search and Rescue (SAR) operations by enabling faster data collection over large or difficult-to-access areas. Beyond simply covering ground quickly, drones equipped with cameras and GPS allow for aerial imaging that can be processed for automated object detection and precise geolocation. This ability can lead to significant improvement in locating missing humans or objects. For example, Smith et al. [8] demonstrated that using drones in mountain rescues reduced search times by 72% compared to traditional methods. However, real-world applications still face many challenges. Lin et al. [9] noted that regulatory restrictions, such as limitations to restrict to fly beyond visual line-of-sight of the controller pilot flights in 78% of countries, and short battery life usually under 30 minutes for most commercial drones. These problems can hinder prolonged or large-scale operations.

These limitations highlight the need for efficient detection algorithms that can perform accurately and quickly within the drone's operational constraints.

2.2 The Human Detection Challenge

Detecting humans from aerial imagery remains fraught with difficulties due to factors such as occlusion, scale variance, and false positives. Occlusion: Rasmussen et al. [10] found that, in forests, trees and bushes obscure 42% of targets in from the drone's camera, while Brown et al. [11] showed thermal camera doesn't work well if the plants cover the body heat signatures. Scale Variance: A person's size in drone imagery shrinks drastically with altitude. Wang and Ng [12] study shown that a 0.5meters tall human appears as just 5 pixels at 100m altitude in the image, making the detection algorithms are very difficult to detect. Wang et al. [13] explicitly warns against relying on targets smaller than 10 pixels. False Positives: The NATO STO [14] reported that 35% of detections in testing areas were animals, shadows, or trash. Leading to a critical issue when lives are at stake.

In the past decade, object detection has improved a lot because of the development in deep learning. Among these, the You Only Look Once (YOLO) framework, initially proposed by Redmon et al. [15], has gained prominence due to its ability to perform real-time object detection with high accuracy. Unlike traditional two-stage detectors like R-CNN, YOLO only needs one step, directly predicting bounding boxes and class probabilities from the input image in one evaluation pass. This architecture drastically reduces inference time, making it suitable for time-sensitive applications such as autonomous driving cars, UAV surveillance, and real-time video analysis. The YOLO algorithm is still continuously developed over the years, including YOLOv3 [16], YOLOv8, and YOLOv11, which have introduced architectural refinements, improved backbone networks, and enhanced training strategies to achieve better trade-offs between accuracy and computational efficiency. These new models are very helpful for SAR tasks, where speed and accuracy are both important. Table 2.1 shows how these models perform.

2.3 Photogrammetry

The geolocation of detected objects is derived through photogrammetric transformations using the Universal Transverse Mercator (UTM) projection system with the WGS84 reference ellipsoid. This method ensures compatibility with global mapping standards by converting pixel coordinates from drone imagery into real-world geographic positions. A key factor in this process is the Ground Sampling Distance **(GSD)**, which quantifies the ground area represented by a single pixel [17]. Calculated as:

$$GSD = \frac{S_w \cdot h}{f \cdot N}$$

where S_w is the camera sensor width, h is the drone altitude, f is the focal length, and N is the image width in pixels. Smaller GSD values (achieved at lower altitudes) enhance detection accuracy by increasing pixel density, while higher altitudes expand coverage at the expense of resolution.

For nadir (vertically downward) camera angles, distortion is minimized, simplifying object detection and geolocation. However, tilted camera angles introduce geometric distortions that require corrections for drone orientation—roll (θ, lateral tilt), pitch (α, forward/backward tilt), and yaw (∅, rotation around the vertical axis). These corrections are applied using rotation matrices [18]:

$$R_z(\emptyset) = \begin{bmatrix} cos\,\emptyset & -\sin\emptyset & 0 \\ \sin\emptyset & \cos\emptyset & 0 \\ 0 & 0 & 1 \end{bmatrix}, R_y(\alpha) = \begin{bmatrix} cos\,\alpha & 0 & \sin\alpha \\ 0 & 1 & 0 \\ -\sin\alpha & 0 & \cos\alpha \end{bmatrix}, R_x(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix}$$

The combined rotation matrix $R_I^G = R_z(\emptyset) \cdot R_y(\alpha) \cdot R_x(\theta)$ aligns the camera's local coordinate system with the global UTM coordinates. Then, pixel coordinates are mapped to real-world locations using the camera matrix (K) and scaled by $\lambda = \frac{h}{f \cdot cos\,\theta \cdot cos\,\alpha}$ to correct perspective distortion.

The final geolocation combines scaled rotations with the drone's GPS position:

$$\overrightarrow{CH}^{w} = \lambda \cdot R_I^G \left(\overrightarrow{CH}^{i}\right)_{normalized} + \varepsilon$$

where $\overrightarrow{CH}^{w}$ represents the object's global coordinates, $\overrightarrow{CH}^{i}_{normalized}$ is the normalized image vector, and ε accounts for residual errors (e.g., lens distortion, GPS noise).

The derived local coordinates are converted to WGS84 latitude/longitude using Coordinate Reference System (CRS) transformations [19], ensuring interoperability with mapping tools like Leaflet. This framework balances technical precision with real-world scalability, addressing the operational trade-off between detection accuracy and coverage efficiency.

2.4 The Measurement of Accuracy Metrics

In evaluating the performance of a classification model, particularly in supervised machine learning tasks, a confusion matrix serves as a fundamental analytical tool [20]. It presents a tabulated comparison between the predicted labels output by the model and the actual ground-truth labels, offering a comprehensive view of classification accuracy. A fundamental tool for evaluating classification performance is the confusion matrix, which categorizes predictions as:

- True Positives (TP): Instances correctly predicted as belonging to the positive class.
- True Negatives (TN): Instances correctly predicted as belonging to the negative class.
- False Positives (FP): Instances incorrectly predicted as positive (Type I error).
- False Negatives (FN): Instances incorrectly predicted as negative (Type II error).

The two most common metrics derived from this matrix are:

- Precision (positive predictive value) measures how accurate the detection results are. Precision is a good metric for the model if the cost of a False Positive is high.
- Recall (sensitivity) quantifies the number of positive classes predicted from all positive examples. Recall is a good measure when the cost of a False Negative is high.

In object detection tasks, model performance is frequently evaluated using Intersection over Union (IoU) and Average Precision (AP) metrics. These metrics provide insight into both the localization and classification accuracy of predicted bounding boxes. Intersection Over Union (IoU) is used to find the difference between the ground truth annotation and the predicted bounding boxes

Average Precision (AP) summarizes the precision-recall curve, which is obtained by varying the confidence threshold of detections. Precision is defined as the ratio of true positive detections to the total number of predicted positives, while Recall measures the proportion of true positives identified out of all actual instances. The Mean Average Precision (mAP) metric integrates both precision and recall by averaging the area under the precision-recall curve across various confidence thresholds and object classes. Two specific versions are:

- mAP@0.5: average precision at IoU threshold of 0.5.
- mAP@0.5:0.95: average precision across IoU thresholds from 0.5 to 0.95 in increments of 0.05.

2.5 The Measurement of Geolocation Error

To measure the geolocation error distance, the Haversine formula is used to calculate the great circle distance between two points on the surface of a sphere. The parameter used to calculate the distance is the radius and its latitudinal and longitudinal coordinates. To calculate the spheroid object like the Earth, the radius used for the formula is the mean radius distance which will give some erroneous but still often acceptable for many geospatial navigation applications [21]. Given two points on the globe, defined by their latitude and longitude in radians, the Haversine formula calculates the distance between these points.

3 Data and Methodology

3.1 Overview of the Framework Design

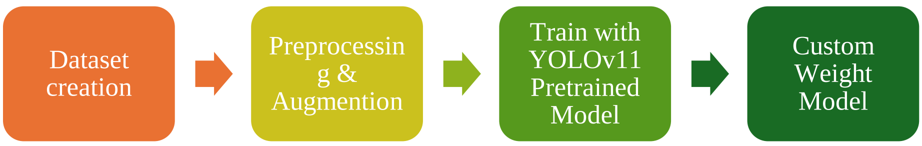


Fig. 1. The framework used for train custom model with YOLOv11

Custom Model Training Framework: This pipeline trains a YOLOv11-based object detection model using drone-captured imagery. The process begins with dataset creation, where clothing items are labeled as human proxies in YOLO format across diverse altitudes (5–30m) and backgrounds (grass, soil, concrete). Raw images undergo preprocessing (resizing to 640×640 pixels) and augmentation (rotation, scaling, brightness adjustments) to enhance robustness. Transfer learning is applied using pretrained COCO dataset weights, enabling efficient fine-tuning on the custom dataset. The output is a set of custom weights optimized for detecting small objects in aerial imagery, balancing speed and accuracy for SAR applications.

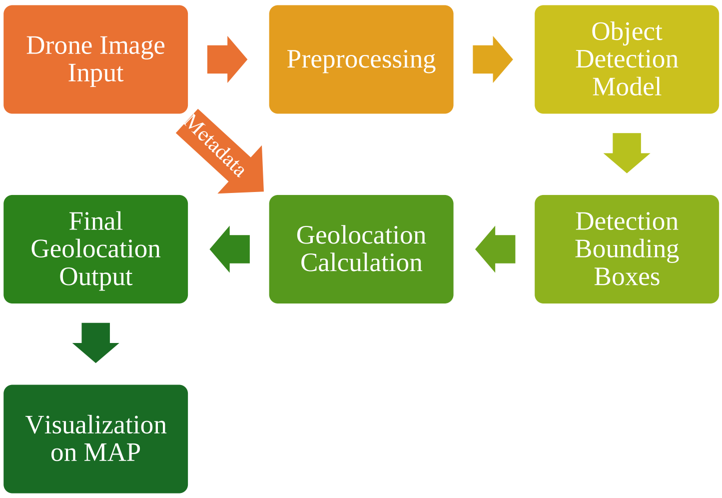


Fig. 2. The framework used for train custom model with YOLOv11

Detection and Geolocation Framework: Deployed in real-time SAR missions, this module processes drone imagery to detect objects and compute their geographic coordinates. Input images, embedded with metadata (GPS, altitude, camera parameters), are preprocessed to match the trained model's input dimensions (640×640). The YOLOv11 model generates bounding boxes around detected objects, whose pixel coordinates are converted to real-world locations using photogrammetric transformations. This involves compensating for drone orientation (yaw, pitch, roll), lens distortion, and ground sampling distance (GSD). Outputs include annotated images displaying detected coordinates, enabling rapid spatial validation by rescue teams.

The frameworks are designed to address the SAR-specific trade-off between detection accuracy and operational speed. By integrating altitude-aware training and geolocation error correction, the system optimizes performance across varying flight conditions, ensuring scalability for real-world deployment.

3.2 Dataset Design

In this independent study, human detection is simulated using clothing to represent a human. This approach is adopted due to various constraints, such as ethical concerns and limitations in data collection. In addition, open-source datasets lack human images captured from various altitudes that range from low to high altitudes. Moreover, for effective model training, data collected across diverse backgrounds and time periods. In this

Independent study, the total use of clothing for the training model is a total of 20, consisting of long-sleeved shirts, sleeveless shirts, t-shirts, shorts, and trousers. The color and pattern selection for model training is also a combination of dark to bright colors, plain to camouflage patterns. All the dataset is collected and labeled into a single class as “clothing" for training aiming to simplify the detection model and reduce the number of needed datasets.

3.3 Dataset collection

The data collecting part uses three different backgrounds of the object for the model including grass, soil, and concrete as shown in Table 1. The height of the collecting drone image varies from a height above ground of 5m up to 30m. Most of the image collection is in a range of 5m and 20m to avoid training with very small objects since YOLO is used to train with sizes of 640x640 pixels.

Table 1. Total clothing class labeled for training model

Background	Height				
	5m	10m	15m	20m	>20m
Soil	40	74	88	168	248
Concrete	56	72	172	183	158
Grass	96	137	161	209	55
Total	192	283	421	560	461

The total image for training the model is as shown in Table 2.

Table 2. Total image in the dataset

Number of training objects in each image	Number of images
0	22
4	280
5-10	7
10-15	9
>15	33
Total	351

In the model training process, the image is split to use as training and validation as 84% and 16% which gives a total number of training 295 and 56 validation images respectively.

3.4 Dataset Preprocessing and Augmentation

After collecting the dataset, all the images are annotated to YOLOv11 format then resized to 640 and augmented with the parameter by using https://roboflow.com [22] as shown in Table 3.

Table 3. Augmentations parameter

Augmentation	Parameter
90° Rotate	Clockwise, Counter-Clockwise
Crop	0% Minimum Zoom, 20% Maximum Zoom
Rotation	Between -15° and +15°
Shear	±10° Horizontal, ±10° Vertical
Saturation	Between -15° and +15°
Hue	Between -25% and +25%
Brightness	Between -15% and +15%
Exposure	Between -10% and +10%
Blur	Up to 2.5px
Noise	Up to 0.1% of pixels

The augmentation applies to training for the model's robustness with real-world scenarios. The number of augmented images is multiplied by 3 (including one of the original image), giving a total of training images after an augmented total of 885 images.

3.5 Model Training

The model was trained using a pretrained YOLOv11 model (yolov11l.pt), which was fine-tuned on the custom dataset. The training utilized standard YOLO hyperparameters and was implemented in Python 3.13.3 using Ultralytics' YOLOv11 framework built on PyTorch.

To explore the effects of image resolution on detection accuracy and computational cost, the model was tested on input sizes ranging from 640×640 up to 5472×3648, including 720×480, 1080×720, 1620×1080, 2432×1620, 3648×2432 to maintain the original 3:2 aspect ratio.

3.6 Drone Flight and Testing

The search method of the flight is using exhaustive search in the testing area. The starting point of each flight height is determined by the DJI Ground Station Pro (DJI GS Pro) algorithm. Each drone flight followed a pre-programmed search path created using DJI GS Pro, with the drone capturing images in nadir angle (camera pointing directly downward) to cover the testing area of 10,000 m². The search path was designed in a grid pattern with 75% front overlap and 60% side overlap to make sure the area is fully covered and has enough overlap between each image. Hover & Capture mode was used to capture the images at each point while keeping the camera facing downward.

The testing experiment was done at four different flight altitudes: 20 meters, 40 meters, 60 meters, and 80 meters above ground level. For each altitude, the flight time and number of images captured were recorded. These images were used for testing with the trained YOLOv11 model by using validation mode. The model outputs detection results including bounding boxes and the class label.

Each detection output was processed to calculate the geolocation coordinates of the object. The center point of each bounding box was used and combined with metadata from the drone such as GPS location, altitude, and camera parameters. Then the real-world latitude and longitude were calculated using the geolocation transformation process.

This testing was designed to evaluate the model's performance in different flight heights. The performance metrics used in the evaluation are mAP@0.5, mAP@0.5:0.95, precision, recall, and detection time. The final detection results also include the bounding boxes and geolocation coordinates of each detected object.

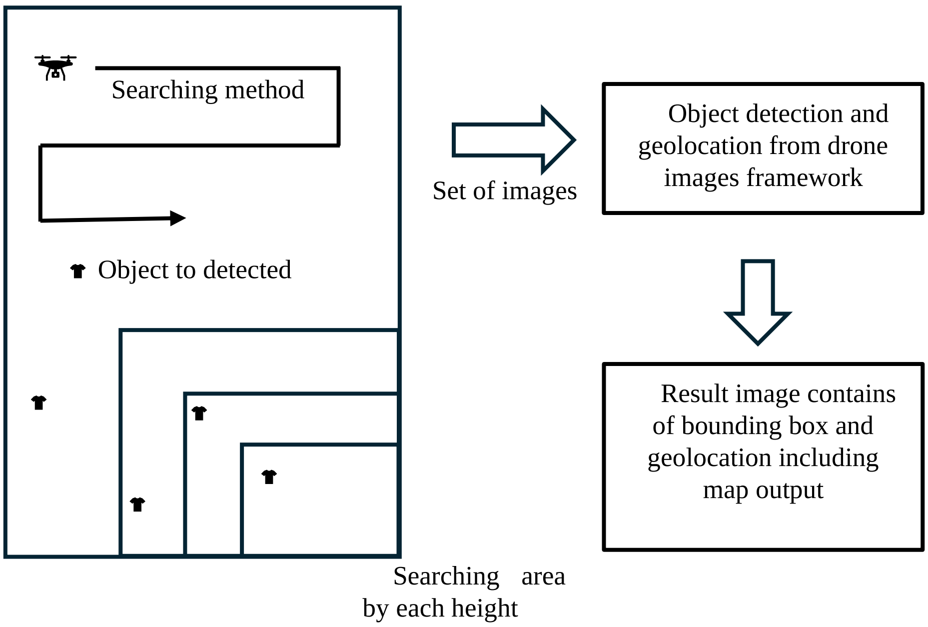


Fig. 3. Design of drone flight path in testing of framework

3.7 Evaluation Metrics

The system's performance is evaluated using metrics that assess detection accuracy, geolocation precision, and operational efficiency, ensuring alignment with search and rescue (SAR) mission requirements.

Detection Accuracy Metrics

- Mean Average Precision (mAP):
    - mAP@0.5: Measures detection accuracy at an Intersection over Union (IoU) threshold of 0.5, reflecting real-world SAR needs where approximate localization suffices.
    - mAP@0.5:0.95: Evaluates performance across IoU thresholds from 0.5 to 0.95, providing a stricter assessment of bounding box precision.
- Precision and Recall:
    - Precision quantifies the ratio of true positive detections to all positive predictions, minimizing false alarms.
    - Recall measures the model's ability to identify all true positives, critical for minimizing missed targets in SAR operations.

Geolocation Accuracy

- Haversine Distance: Computes the great-circle distance between predicted and ground-truth coordinates using latitude/longitude values

$$d = 2R\, arcsin\left(\sqrt{sin^2\left(\frac{\emptyset_2 - \emptyset_1}{2}\right) + cos(\emptyset_1)\, cos(\emptyset_2)\, sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)}\right)$$

where R is Earth's radius (6,371 km), $\emptyset_2 - \emptyset_1$ and $\lambda_2 - \lambda_1$ are latitude and longitude differences are in radians.

Operational Efficiency Metrics

- Computation Time: Total processing time per image, measured in seconds, to ensure real-time applicability.
- Efficiency Score: Balances detection accuracy (mAP@0.5) against computation time:

$$Efficiency\ Score = \frac{Mean\ Average\ Pricision}{Computation\ time}$$

- Missing Rate: Complement of Recall $(1 - Recall)$, quantifying undetected targets.

3.8 Tools and Hardware

- Drone: DJI Phantom 4 Pro V2.0, equipped with a 1-inch CMOS sensor (focal length: 8.8 mm, 20MP, JPEG format, image size: 5472×3648) [23]
- GPS: Garmin GPSMAP® 67i (Multi-band GNSS) [24]
- Hardware: NVIDIA RTX 2070 SUPER, AMD Ryzen 7 3700X, 80 GB RAM

- Libraries: PyTorch, OpenCV, Ultralytics YOLO, NumPy, Pandas, Matplotlib, Folium, Pyproj
- Software: DJI GS Pro (for flight path planning), Roboflow (for image annotation and augmentation), Exiftool (for metadata extraction)
- Development Environment: Python 3.13.3 on Windows 11

4 Results

4.1 Experimental Setup Summary

The testing of the proposed framework was performed in a real-world environment at Thanawan Park, Chiang Mai, over a fixed area of 10,000 m^2. The drone was deployed at four different altitudes: 20 m, 40 m, 60 m, and 80 m above ground level. The testing at each altitude followed the setup described in Section 3.6, using a grid search path generated by DJI Ground Station Pro (GS Pro) with 75% front overlap and 60% side overlap.

Images captured from each altitude were used as input to the trained YOLOv11 model in validation mode. The model returned detection outputs including bounding boxes and object classes. For each detection, geolocation coordinates were calculated based on drone metadata and image geometry.

This setup allowed the study to evaluate the system's detection accuracy, geolocation precision, and performance trade-offs at different flight heights. The following sections present the analysis and comparison of these results across altitudes and image resolutions.

4.2 Area Coverage and Search Duration

The drone was flown at 4 different altitudes over the same area (fixed at 10,000 m^2) using the same flight plan. The data collected in terms of the coverage area compared to the time of searching drone is shown in Table 4.

Table 4. The flight coverage area of each flight altitude

Altitude (m)	Flight Time (min)	Area per Minute (m^2/min)	Seconds per m^2
20	13.3	750.9	0.08
40	4.3	2316.6	0.03
60	2.6	4347.8	0.014
80	1.2	8108.1	0.0074

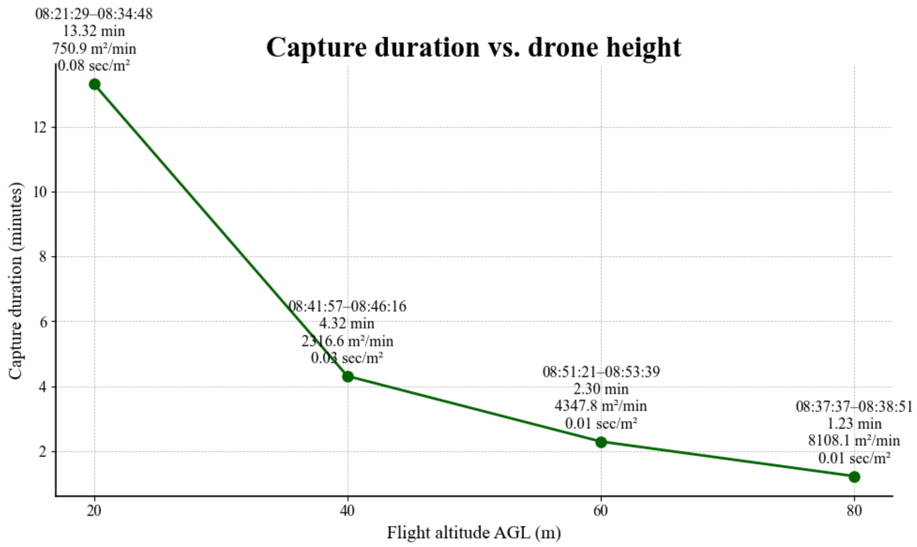


Fig. 4. Comparison between drone searching time and drone flight altitude

Figure 4 presents the searching time and area coverage across four flight altitudes (20 m, 40 m, 60 m, and 80 m). In this study, the search area was fixed at 10,000 m². It was observed that flights conducted at lower altitudes required longer hovering times to capture images, resulting in increased overall search duration compared to higher-altitude flights. As illustrated in Figure 4.5, increasing flight altitude significantly reduces the total search time. This is attributed to the larger ground coverage per image at higher altitudes, which exponentially decreases the required image capture frequency and duration.

4.3 Geolocation Accuracy

After detection, the geolocation of the object is calculated and compared with the real location (measured by GNSS). The difference is measured by the Haversine distance formula. The results are shown in Table 5.

Table 5. Average of haversine error at a different flight altitude

Altitude (m)	Avg Haversine Error (m)
20	9.64
40	23.63
60	24.37
80	32.79

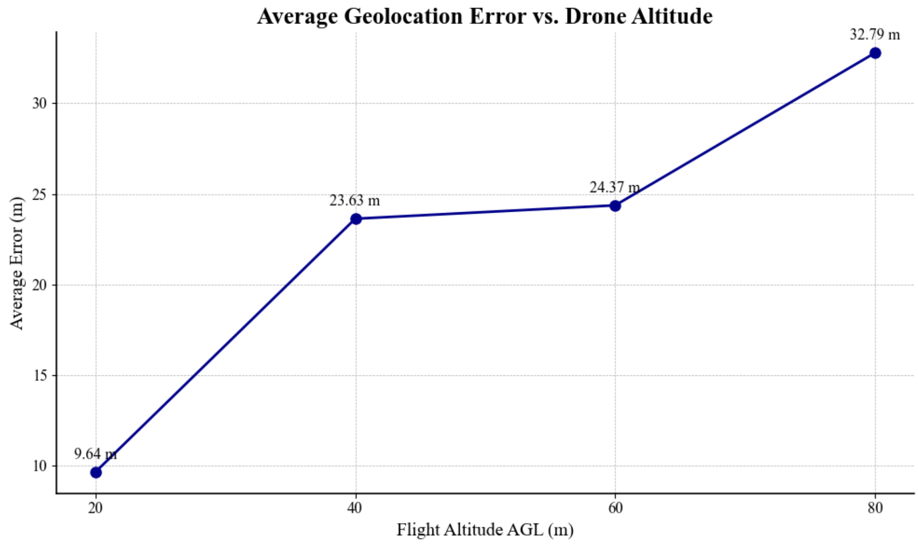


Fig. 5. Comparison between drone geological error and flight altitude

Figure 5 presents the average of Haversine error across four flight altitudes (20 m, 40 m, 60 m, and 80 m). The increase in error with flight altitude is what the study expected due to several reasons. First, the higher image from drones presents the object presented by fewer pixels. This led to ground sampling distance increases with this distance. Moreover, the error does not increase in direct proportion to altitude because with increasing altitude and the geolocation of the object image depends on its position within the image. Objects closer to the image center tend to have more accurate geolocation while those farther off center are more affected by lens distortion and errors in pixel-to-coordinate conversion increase. Furthermore, any pixel-level errors translate into larger errors in larger real-world coordinates also become greater.

4.4 Detection Accuracy Across Image Resolutions and Altitudes

The custom YOLOv11 model was tested with various image input resolutions. The mAP@0.5, mAP@0.5:0.95, precision, recall, and computation time were recorded. The results of the testing are shown in Table 6.

Table 6. Accuracy of the model with custom-trained weight adjust by input resolution

Resolution (px)	Altitude	mAP@0.5	mAP@0.5: 0.95	Precision	Recall	Time (s)
640 × 640	20 m	0.375	0.260	0.458	0.500	42.17
	40 m	0.156	0.059	0.112	0.519	22.24
	60 m	0.012	0.005	0.056	0.054	6.70
	80 m	0.000	0.000	0.000	0.000	6.22
720 × 480	20 m	0.390	0.275	0.582	0.450	29.33

	40 m	0.221	0.118	0.323	0.407	8.66
	60 m	0.086	0.042	0.163	0.216	7.33
	80 m	0.004	0.001	0.009	0.021	6.91
1080 × 720	20 m	0.327	0.258	0.595	0.442	29.59
	40 m	0.326	0.220	0.377	0.696	8.92
	60 m	0.152	0.086	0.253	0.351	8.19
	80 m	0.030	0.017	0.099	0.128	7.29
1620 × 1080	20 m	0.306	0.246	0.453	0.450	60.59
	40 m	0.375	0.272	0.527	0.575	10.56
	60 m	0.193	0.139	0.429	0.324	8.59
	80 m	0.064	0.040	0.132	0.255	7.55
2432 × 1620	20 m	0.279	0.242	0.352	0.500	147.55
	40 m	0.362	0.276	0.425	0.593	64.43
	60 m	0.210	0.161	0.374	0.339	33.88
	80 m	0.122	0.090	0.281	0.199	11.84
3648 × 2432	20 m	-(1)	-(1)	-(1)	-(1)	-(1)
	40 m	-(1)	-(1)	-(1)	-(1)	-(1)
	60 m	0.298	0.232	0.332	0.432	82.42
	80 m	0.205	0.157	0.378	0.340	44.43
5472 × 3648	all	-(1)	-(1)	-(1)	-(1)	-(1)

Note (1): The GPU runs out of memory while testing the model with the image resolution input

4.5 Comparison of Flight Altitudes Across Image Resolutions and Detection Accuracy Metrics

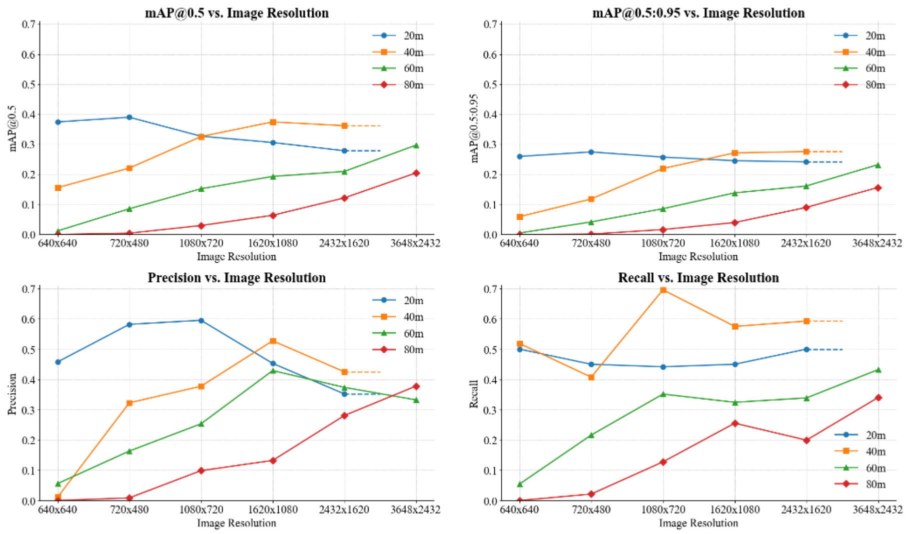


Fig. 6. The relationship between accuracy metrics and image resolution in different flight altitudes

Figure 6 presents the effect of image resolution on detection performance metrics such as Mean Average Precision (mAP), Precision, and Recall across four flight altitudes (20 m, 40 m, 60 m, and 80 m).

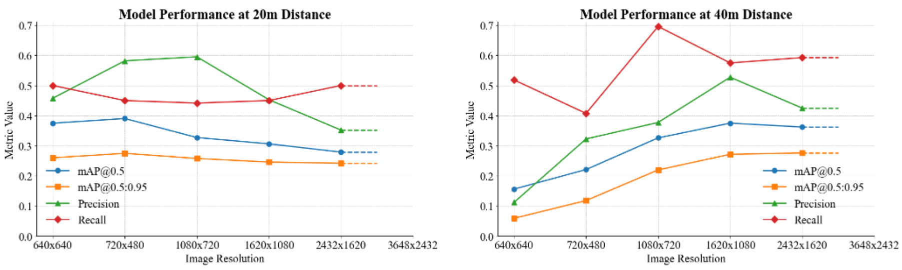
Mean Average Precision (mAP) varies distinctly with altitude and resolution. At the lowest altitude (20 m), Mean Average Precision starts relatively high, even at low resolutions, slightly increasing at 780×480 resolution before experiencing a minor decline at the highest resolution tested. In contrast, at 40 m, Mean Average Precision improves substantially with increasing resolution, peaking at 1620×1080 pixels, followed by a slight drop at the highest resolution. At 60 m and 80 m, detection performance begins lower but steadily improves with increasing resolution, highlighting the growing importance of high image detail for accurate detection at greater heights.

Precision trends echo similar altitude-dependent patterns. At 20 m, Precision peaks at a medium resolution (1080×720) and then decreases at higher resolutions, potentially due to overfitting or noise, as objects are already well-resolved. For 40 m and 60 m, Precision increases with resolution, reaching maximum values at 1620×1080 pixels before slightly declining at the highest tested resolution. At 80 m, Precision starts near zero at low resolution but rises steadily, surpassing the 60 m level at the maximum resolution of 3648×2432 pixels, underscoring the critical role of high-resolution imaging for high-altitude detection.

Recall performance also demonstrates altitude-specific trends. At 20 m, Recall remains consistently stable across all resolutions, indicating that even low-resolution images suffice for reliable detection at close range. At 40 m, Recall peaks at 1080×720 pixels but declines at higher resolutions, possibly due to image complexity or noise effects. Both 60 m and 80 m altitudes show a steady increase in Recall with resolution, with the highest values observed at 3648×2432 pixels, reflecting the enhanced capability to detect smaller, less distinct objects when image detail improves.

These results collectively demonstrate that the influence of image resolution on detection accuracy is strongly dependent on flight altitude. Lower altitudes allow effective detection even at moderate resolutions due to a larger apparent object size. In contrast, higher altitudes necessitate increased image resolution to compensate for reduced object size and detail, thereby improving detection metrics significantly.

4.6 Comparison of Flight Time and Computation Time at Each Altitude



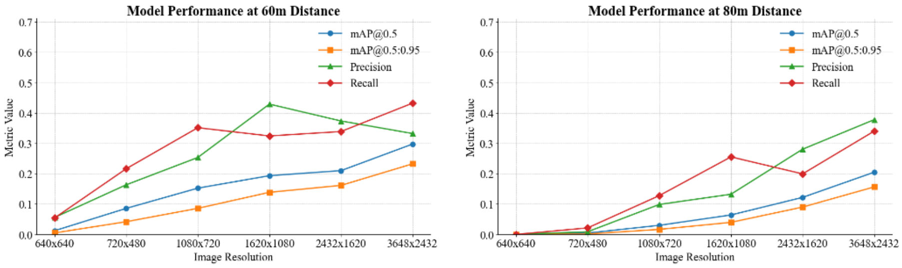


Fig. 7. Comparison between accuracy metrics across different image resolutions in different flight altitudes

Figure 7 presents evaluation metrics—including mAP@0.5, mAP@0.5:0.95, Precision, and Recall—analyzed across varying image resolutions at four flight altitudes (20 m, 40 m, 60 m, and 80 m).

At 20 meters, both mAP@0.5 and mAP@0.5:0.95 exhibit a decreasing trend as image resolution increases. The highest values are observed at the lowest resolution (720×480), with a gradual decline up to 2432×1620 pixels. Precision peaks at a medium resolution (1080×720), nearing 0.6, before steadily decreasing at higher resolutions, likely due to increased noise or overfitting since objects are already clearly visible at close range. In contrast, Recall remains relatively stable around 0.45 to 0.5 across all resolutions, indicating that while overall detection quality is sensitive to resolution, the model's ability to detect objects remains consistent. These results suggest that for low-altitude flights, low to medium resolutions optimize detection performance, and higher resolutions may reduce efficiency without significant gains.

At 40 meters, the metrics show more complex trends. Both mAP@0.5 and mAP@0.5:0.95 increase with resolution, peaking at 2432×1620 pixels before stabilizing. Precision starts very low but rises consistently, reaching a peak at 1620×1080 pixels, followed by a slight decline at the highest resolution. Recall is highest among all altitudes at lower resolutions, peaking at 1080×720 before declining and stabilizing at higher resolutions. This indicates that moderate to high resolutions improve detection accuracy at 40 m, but there is a trade-off, as improvements in Precision and mAP come at the cost of decreased Recall beyond certain resolution thresholds. This trade-off highlights the importance of balancing metrics rather than optimizing a single one.

At 60 meters, all metrics generally improve as resolution increases. Both mAP@0.5 and mAP@0.5:0.95 show consistent upward trends. Precision peaks at 1620×1080 pixels before slightly decreasing at the highest resolution, while Recall sharply rises from lower resolutions, peaks at 1080×720, dips slightly, then increases again at higher resolutions. This pattern suggests that increasing resolution substantially enhances detection accuracy at 60 m, although beyond a certain resolution, gains in Precision plateau or slightly decline, indicating diminishing returns at ultra-high resolutions.

At 80 meters, evaluation metrics demonstrate continuous improvement with increasing resolution. Both mAP metrics steadily increase without plateauing or declining at the highest tested resolution, indicating a strong positive correlation between resolution and detection accuracy at extended distances. Precision rises sharply as resolution improves, and Recall, after an initial increase and slight dip, ultimately peaks at the highest resolution. Unlike lower altitudes, there is no observable degradation in performance at ultra-high resolutions, underscoring the necessity of high-resolution imagery for effective detection at greater flight heights.

4.7 Efficiency Score

For the trade-off between the model accuracy and time. Efficiency Scores are created to evaluate model efficiency, which is very important when working with many images in real time. Even though flight time is the largest part of total time at each height, we use computation time to measure efficiency because it shows how well the detection system performs. These scores help us see how good the model is compared to the time it takes to make predictions. A higher score means the model gives better accuracy with less processing time. This is very useful in real-time or low-resource situations. The Efficiency Score across different altitudes in each image resolution is shown in Table 7.

Table 7. Efficiency Score of each flight altitude and image resolution

Resolution	20m	40m	60m	80m
640x640	0.00889	0.00702	0.00181	0.00000
720x480	0.01331	0.02552	0.01169	0.00064
1080x720	0.01106	0.03656	0.01860	0.00410
1620x1080	0.00506	0.03551	0.02251	0.00848
2432x1620	0.00189	0.00562	0.00620	0.01030
3648x2432	N/A	N/A	0.00361	0.00461

From Table 7, the highest Efficiency Score of the model is achieved at the flight altitude of 40m. At the resolution of 1080x720, giving the value of 0.03656.

4.8 Missing Rate

In the context of evaluating detection effectiveness, it’s equally important to assess how many true positives were missed by the model, which is referred to as the Missing Rate. Since the Recall is the ratio of true positives to all actual positives, and vice versa, the Missing Rate is simply the complement of Recall. The Missing Rate across different altitudes in each image resolution is shown in Table 8.

Table 8. Missing Rate of each flight altitude and image resolution

Resolution	20m	40m	60m	80m
640x640	0.5000	0.4815	0.9459	1.0000
720x480	0.5500	0.5926	0.7838	0.9787
1080x720	0.5584	0.3041	0.6486	0.8723
1620x1080	0.5500	0.4251	0.6757	0.7447

2432x1620	0.5000	0.4074	0.6612	0.8007
3648x2432	N/A	N/A	0.5676	0.6596

From Table 8, the lower Missing Rate better the model performance. Note that the lowest Missing Rate from altitude is 0.5000 in both resolutions of 640x640 and 2432x1620 pixels. The lower resolution is considered more appropriate since it requires less computation time. The lowest Missing Rate of the model is achieved at the flight altitude of 40m. At the resolution of 1080x720, giving the value of 0.3041.

5 Discussion and Conclusion

5.1 Efficiency Trade-off: Accuracy vs. Time vs. Area Coverage

In real-world SAR operations, not only is the objective of each mission different, but the constraint of each individual operation also varies significantly. Factors such as the size of the search area, battery limitations, and onboard computational capacity must all be considered when planning and executing detection missions.

- If Accuracy is the Priority: Since the mAP@0.5 and mAP@0.5:0.95 are both the same measurement, the mAP@0.5 is the only represented here. As shown in Table 9, the comparison of drone performance across different altitudes by evaluating three competing factors: Mean Average Precision (mAP@0.5), processing time, and coverage efficiency. Mean Average Precision is prioritized when accuracy is the primary concern because it reflects Precision across different levels of Recall and considers both false positives and false negatives.

Table 9. Comparison between accuracy vs. time vs. area coverage

Height (m)	Optimal resolution	Best mAP@0.5	Computation time(s)	Coverage area per minute (m^2/min)	Overall Efficiency
20	720x480	0.3904	~29	750.9	Best accuracy, moderate computation time
40	1620x1080	0.3750	~11	2316.6	High accuracy, lowest computation time
60	3648x2432	0.2976	~82	4347.8	Normal accuracy, highest computation time
80	3648x2432	0.2050	~44	8108.1	Lowest accuracy, high computation time

From the Table 9, the highest mAP@0.5 is achieved at the lowest testing flight altitude of 20 meters at 720x480 pixels resolution.

- If Balance of Accuracy Needed: The balance of the model between accuracy and time can be determined using the Efficiency Score, which evaluates both accuracy and time. In Table 10, it compares drone performance across different altitudes by evaluating three competing factors: Efficiency Score, processing time, and coverage efficiency.

Table 20. Comparison between Efficiency Score vs. time vs. area coverage

Height (m)	**Optimal resolution**	**Best Efficiency Score**	**Computation time(s)**	**Coverage area per minute (m^2/min)**	**Overall Efficiency**
20	720x480	0.01331	~29	750.9	Low efficiency, low area/min
40	1080x720	0.03656	~9	2316.6	Best efficiency, moderate area/min
60	1620x1080	0.02251	~9	4347.8	Normal efficiency, high area/min
80	2432x1620	0.01030	~12	8108.1	Lowest efficiency, highest area/min

From the Table 10, the best efficiency of the model is achieved at a flight altitude of 40 meters at a 1080x720 pixels resolution.

- If Completeness of Detection is the Priority: Sometimes in critical SAR operations, the priority of the mission requires completeness of detection to be more important than accuracy or model efficiency. From Table 11 compares drone performance across different altitudes by evaluating three competing factors: Missing Rate, processing time, and coverage efficiency.

Table 31. Comparison between Missing Rate vs. time vs. area coverage

Height (m)	**Optimal resolution**	**Lowest Missing Rate**	**Computation time(s)**	**Coverage area per minute (m^2/min)**	**Overall Efficiency**
20	640x640	0.5000	~42	750.9	Moderate performance, high computation time
40	1080x720	0.3041	~9	2316.6	Best performance, lowest computation time

60	3648x2432	0.5676	~82	4347.8	Low performance, highest computation time
80	3648x2432	0.6596	~44	8108.1	Lowest performance, high computation time

From the Table 11, the lowest Missing Rate of the model is achieved at a flight altitude of 40 meters at a 1080x720 pixels resolution. Therefore, developing a practical guideline requires evaluating performance trade-offs under these constraints to select the most appropriate configuration for a given operational scenario.

5.2 Conclusions

From the human detection and geolocation framework of this independent study, which uses clothing as an object for detection. The result confirmed that increasing drone flight altitude influences greater errors in both object detection and geolocation, which aligns with expectations due to several factors.

In terms of detection performance, the drone is highly dependent on the flight altitude. The object needed pixel size needed to identify generally increases along the altitude. The observation also shows that each drone altitude trend to have an optimal resolution for detecting objects. Increase images resolution beyond this optimal resolution, the performance of the model surprisingly does not increase but reduces the performance of the model. At higher altitudes, larger image resolutions are necessary to maintain reasonable accuracy. However, there is a trade-off that even higher resolution cannot fully compensate for the loss in performance such as seen at an altitude of 80 meters. Moreover, the increase in image resolution for object detection sharply raises the computation time. The impact of these effects makes the great efficiency of model.

In contrast, as flight altitude increases, the coverage area per time is greater, as seen in the study the altitude of 80 meters covers the area faster than 4 times compared to the altitude of 20 meters. Therefore, efficient model deployments must account for various constraints such as the accuracy needed, computation resources, latency requirements, etc.

The computation time also shows a notable trade-off between high and low resolutions. At higher altitudes, the border images from drones are bigger, resulting in fewer images needed for model prediction. Vice versa, at lower altitudes, the number of images required to process is larger than images from higher altitudes. This presents a significant change in balancing detection accuracy and computation resources, especially in real-time detection.

In the geolocation extraction framework shows that not only flight altitude that affected the error but also the location of the object in the image plane. As the altitude increases, the objects are represented with fewer pixels leading to a larger ground sampling distance. Furthermore, any pixel-level errors where used to translate into real-world coordinates will become larger.

The findings guidelines in this study are introduced with an intention to balance the altitude, resolution, performance, computation time, and coverage area. If needed of model accuracy metrics, the flight altitude should be low (20 meters in this study). The high flight altitudes are suitable for large area operations but also require larger image resolutions with high available computation costs to compensate for the performance needed. In general operation with no specific constraint or requirement, the 40 meters with a resolution of 1080x720 pixels are recommended.

5.3 Discussion

In this study, the YOLOv11 object detection model was used to identify and extract the geolocation using an image captured by a drone at a different flight altitude and computed accuracy in varied resolutions. While YOLO is widely recognized for its real-time detection capabilities and computational efficiency, the results from this study reveal notable limitations in accuracy, particularly as the flight altitude increases. Specifically, as the altitude increases, the model's performance in both the detection and geolocation parts becomes greatly reduced.

The object detection model in this study uses augmentation methods with an image in three backgrounds. The addition of a real-world visual complex environment such as foliage, cluttered backgrounds with debris or disaster-specific textures, and more varied terrains such as mud or water could help improve detection robustness. Incorporating these elements may enhance the ability to detect objects under challenging and realistic conditions commonly encountered in search and rescue operations.

Since the aim of this study is to focus on the effect of drone altitude on the performance of the model, the model training in this study did not involve tuning any parameters. The YOLO model trains with an image size of 640x640, whereas the original image size is in a 3:2 aspect ratio. This can significantly impact the accuracy of the model due to image distortions.

The pretrained weights used in this study are primarily trained on the COCO dataset, which is great for common image classes such as people, animals, and vehicles. This pretrained model may not generalize well with the aerial image plan taken from a drone.

Although flight time is dominating the computation times, this study focuses on detection efficiency, since it directly affects how fast and scalable the system is in real-time use. Unlike flight time, which stays mostly the same for each flight, computation time varies with image resolution and impacts system performance. To evaluate the balance between accuracy and speed, we use Efficiency Scores, which measure how much accuracy the model achieves for the time it takes to process images.

In this study, object locations are taken from the DJI Phantom 4 camera’s EXIF data, which gives important information for the geolocation part. The use of different cameras or sensors such as with nadir might give more stable results with less image distortion and no gimbal tilts. However, an oblique view can be better in areas with many obstacles, as it can show objects that are hidden from above. This creates a trade-off between accurate mapping and better visibility, which should be chosen based on the needs of the search and rescue mission.

References

[1] M. Hassanalian and A. Abdelkefi, “Classifications, applications, and design challenges of drones: A review,” Prog. Aerosp. Sci., vol. 91, pp. 99–131, 2017. [Online]. Available: https://doi.org/10.1016/j.paerosci.2017.04.003

[2] R. R. Murphy et al., “Search and rescue robotics,” in Springer Handbook of Robotics, B. Siciliano and O. Khatib, Eds. Berlin, Germany: Springer, 2008, pp. 1151–1173.

[3] M. Półka, S. Ptak, Ł. Kuziora, and A. Kuczyńska, “The use of unmanned aerial vehicles by urban search and rescue groups,” in Drones - Applications, InTech, 2018. [Online]. Available: https://doi.org/10.5772/intechopen.73320

[4] S. Waharte and N. Trigoni, “Supporting search and rescue operations with UAVs,” in 2010 Int. Conf. Emerg. Secur. Technol., 2010, pp. 31–36. [Online]. Available: https://doi.org/10.1109/EST.2010.31

[5] M. N. Kiyani and M. U. M. Khan, “A prototype of search and rescue robot,” in Proc. 2nd Int. Conf. Robot. Artif. Intell. (ICRAI), 2016, pp. 208–213.

[6] G. Jocher, J. Qiu, and A. Chaurasia, Ultralytics YOLO (Version 8.0.0) [Computer software]. [Online]. Available: https://github.com/ultralytics/ultralytics

[7] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” arXiv preprint arXiv:1506.02640, 2016. [Online]. Available: https://arxiv.org/abs/1506.02640

[8] G. Smith et al., “Operational efficiency of drones in mountain SAR,” IEEE Trans. Emerg. Topics Comput., vol. 9, no. 3, pp. 1452–1464, 2021. [Online]. Available: https://doi.org/10.1109/TETC.2021.3063022

[9] Y. Lin et al., “Regulatory barriers to drone SAR deployment,” J. Field Robot., vol. 40, no. 2, pp. 412–428, 2023. [Online]. Available: https://doi.org/10.1002/rob.22038

[10] L. Rasmussen et al., “Occlusion in forested SAR,” IEEE Trans. Aerosp. Electron. Syst., vol. 58, no. 1, pp. 210–225, 2022. [Online]. Available: https://doi.org/10.1109/TAES.2021.3081571

[11] C. Brown et al., “Thermal imaging limitations in SAR,” IEEE Sensors J., vol. 22, no. 6, pp. 5012–5020, 2022. [Online]. Available: https://doi.org/10.1109/JSEN.2022.3145992

[12] W. Wang and K. Ng, “Scale variance in aerial detection,” IEEE Robot. Autom. Lett., vol. 8, no. 3, pp. 1567–1574, 2023. [Online]. Available: https://doi.org/10.1109/LRA.2023.3245623

[13] J. Wang, W. Yang, H. Guo, R. Zhang, and G.-S. Xia, “Tiny object detection in aerial images,” in Proc. 25th Int. Conf. Pattern Recognit. (ICPR), Milan, Italy, Jan. 2021, pp. 1–8. [Online]. Available: https://doi.org/10.1109/ICPR48806.2021.9413340

[14] NATO Science and Technology Organization (STO), False Positives in Aerial SAR, Tech. Rep. SAS-IST-182, 2023.

[15] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit.

(CVPR), Las Vegas, NV, USA, Jun. 2016, pp. 779–788. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2016/html/Redmon_You_Only_Look_CVPR_2016_paper.html

[16] J. Redmon and A. Farhadi, “YOLOv3: An incremental improvement,” arXiv preprint arXiv:1804.02767, Apr. 2018. [Online]. Available: https://arxiv.org/abs/1804.02767

[17] P. R. Wolf and B.A. Dewitt, Elements of Photogrammetry with Applications in GIS, 4th ed. New York, NY, USA: McGraw-Hill, 2014.

[18] R. Hartley and A. Zisserman, Multiple View Geometry in Computer Vision, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[19] Z. Zhang, “A flexible new technique for camera calibration,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 11, pp. 1330–1334, 2000. [Online]. Available: https://doi.org/10.1109/34.888718

[20] T. Fawcett, “An introduction to ROC analysis,” Pattern Recognit. Lett., vol. 27, no. 8, pp. 861–874, Jun. 2006. [Online]. Available: https://doi.org/10.1016/j.patrec.2005.10.010

[21] J. P. Snyder, Map Projections—A Working Manual, U.S. Geol. Surv. Prof. Paper 1395, U.S. Govt. Print. Off., 1987.

[22] B. Dwyer, J. Nelson, T. Hansen, et al., Roboflow (Version 1.0) [Computer software], 2024. [Online]. Available: https://roboflow.com

[23] J. Abbott, “DJI Phantom 4 Pro V2.0 review,” TechRadar, Sep. 28, 2020. [Online]. Available: https://www.techradar.com/reviews/dji-phantom-4-pro-v20. [Accessed: April 30, 2025].

[24] Garmin, “GPSMAP 67i,” Garmin. [Online]. Available: https://www.garmin.com/products/outdoor/gpsmap-67i/. [Accessed: April 30, 2025].