Person Detection and Geolocation Estimation in UAV Aerial Images: An Experimental Approach

Sasa Sambolek[1][a] and Marina Ivasic-Kos[2][b]
[1]High school Tina Ujevica, Kutina, Croatia
[2]Faculty of Informatics and Digital Technologies, University of Rijeka and Centre for Artificial Intelligence University of Rijeka, Rijeka, Croatia

Keywords: Drone Imagery, Deep Learning, Person Detection, YOLOv8, Search and Rescue.

Abstract: The use of drones in SAR operations has become essential to assist in the search and rescue of a missing or injured person, as it reduces search time and costs, and increases the surveillance area and safety of the rescue team. Detecting people in aerial images is a demanding and tedious task for trained humans as well as for detection algorithms due to variations in pose, occlusion, scale, size, and location where a person may be in the image, as well as poor shooting conditions, poor visibility, blur due to movement and the like. In this paper, the YOLOv8 generic object detection model pre-trained on the COCO dataset is fine-tuned on the customized SARD dataset used to optimize the model for person detection on aerial images of mountainous landscapes, which are captured by drone. Different models of the YOLOv8 family algorithms fine-tuned on the SARD set were experimentally tested and it was shown that the YOLOv8x model achieves the highest

for real-time use in SARD operations. We have tested three geolocation algorithms in real conditions and proposed modification and recommendations for using in SAR missions for determining the geolocation of a person recorded by drone after automatic detection with the YOLOv8x model.

1 INTRODUCTION

Object detection is a key research area within computer vision, focusing on the precise positioning and recognition of various objects in the image (Zou et al., 2023). Despite achieving promising results in ground-level object detection, the task of object detection in aerial images is still a challenge, especially in its application in search and rescue (SAR) operations (Sambolek & Ivasic-Kos, 2021) whose primary objective is to assist as soon as possible to the casualty and save human lives.

SAR is carried out on different terrains such as mountains, rivers, lakes, canyons. The speed of finding a missing person directly affects their chances of survival, so unmanned aerial vehicles (drones) equipped with RGB cameras and sensors are nowadays commonly included in the search missions. The search area is inspected during the flight and offline with the subsequent analysis of the recorded material if the missing person is not found during the online search. In both cases, artificial intelligence can help track down the missing person, however, the automatic detection of victims is still a challenge (Andriluka et al., 2010; Bejiga et al., 2017; Doherty & Rudol, 2007; Geraldes et al., 2019; Shakhatreh et al., 2019; Sun et al., 2016). When analyzing the recorded material, it is crucial not only to detect the person in the images, but also to estimate the distance of the person from the drone and to geolocate it so that a SAR mission can be organized accordingly.

The primary goal of this paper is to evaluate the effectiveness of the latest version of the widely used YOLO object detector, YOLOv8 (Ultralytics, n.d.-c), in detecting people in drone images. Using the publicly available SARD dataset (Sambolek & Ivasic-Kos, 2021) adapted for object detection in SAR, we fine-tuned different models of the Yolov8 family and conducted an in-depth analysis and comparison of drone-captured person detection

[a] https://orcid.org/0000-0002-5287-2041
[b] https://orcid.org/0000-0002-1940-5089

Sambolek, S. and Ivasic-Kos, M.
Person Detection and Geolocation Estimation in UAV Aerial Images: An Experimental Approach.
DOI: 10.5220/0012411600003654

In *Proceedings of the 13th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2024)*, pages 785-792
ISBN: 978-989-758-684-2; ISSN: 2184-4313

performance. In addition, we have built custom SAR-DAG_overflight dataset for developing and testing the algorithm for determining the geolocation of a detected person.

The structure of this paper is as follows: Section 2 provides an overview of previous research related to YOLO object detectors and person geolocation algorithms. The YOLOv8 family of models and the performance achieved after fine-tuning on the customized SARD dataset are described in Section 3, followed by a description of the geolocation algorithms proposed for use in SAR missions. The experimental part of the work and the metrics used are presented in Section 4 along with the results and explanation. The concluding section highlights the main contributions of this paper.

2 RELATED WORKS

For our proposed method of detection and geolocation of persons in SAR missions, the object detector and the geolocation algorithm are key. In the following, we will focus on the review of the state-of-the-art CNN detectors from the YOLO family (Redmon et al., 2016), which are an example of single-stage detectors that constantly achieve top performance in real time, and algorithms for deterministic geolocation.

2.1 YOLO Object Detectors

The most popular and stable version of YOLO, showcasing improved performance with multi-scale prediction frameworks and a deep backbone network, was introduced by Redmon and Farhadi (Redmon & Farhadi, 2018). Bochkovskiy et al. (Bochkovskiy et al., 2020) developed YOLOv4, which featured significant new features, outperforming YOLOv3 in terms of accuracy and speed. (Ultralytics, n.d.-a) introduced YOLOv5, along with a PyTorch-based variant, bringing remarkable improvements. In 2022, the Meituan Vision AI Department unveiled YOLOv6 (Li Chuyi et al., 2022). YOLOv6 features an efficient backbone, RepVGG or CSPStackRep blocks, PAN topology gates, and efficient separate heads with a hybrid channel strategy. The model also employs advanced quantization techniques, including post-training quantization and channel distillation, resulting in faster and more accurate detectors. In July of the same year, YOLOv7 (Chien-Yao Wang, Alexey Bochkovskiy, 2023) outperformed all existing object detectors in terms of speed and accuracy. It follows the same COCO dataset training approach as YOLOv4 but introduces architectural changes and improvements that enhance accuracy without compromising inference speed. The most recent version of the YOLO family released in January 2023 is YOLOv8 (Ultralytics, n.d.-c) designed for speed and precision for various computer vision applications (Ultralytics, n.d.-c). The architecture of YOLOv8 can be divided into two main components: the backbone and the head. The backbone is like the YOLOv5 model and contains the CSPDarknet53 architecture with 53 convolutional layers, but with the change in the building blocks of the C3 module. The module is now called C2f and all outputs from the gate (bottleneck – 3x3 convolutions with residual connections) were chained, while in C3 only the output from the last gate was used. In the neck, the features are connected directly without forcing the same channel dimensions, which reduces the number of parameters and the total size of the tensor. The head of YOLOv8 consists of several convolutional layers, followed by fully connected layers responsible for predicting bounding boxes, objectivity (probability that the bounding box contains an object), and class probabilities for recognized objects. For class probabilities, the softmax function is used, while the output layer uses the sigmoid function as the activation function.

The loss functions used by YOLOv8 for improving detection, especially when working with smaller objects are: CIoU (Complete Intersection over Union) and DFL (Distribution Focal Loss) for bbox-related losses, and binary cross-entropy for classification loss.

YOLOv8 uses an anchor-free model with a decoupled head for independent object detection, classification, and regression processing. This design allows each branch to focus on its task and contributes to improving the overall accuracy of the model.

2.2 Target Geolocation Algorithms

To calculate the geolocation of objects in the image, an algorithm based on the Earth ellipsoid model is usually used, (Leira et al., 2015; Sun et al., 2016; Wang et al., 2017; Zhao et al., 2019) which uses information about the average height, the field of view of the camera, the width and height of the image, the tilt of the camera and the position of the detected point within the image. This algorithm is easy to calculate, but it is not precise because it considers the average elevation information as the reference height for the target, which leads to significant positional inaccuracies, especially in regions with significant topographic relief. Figure 1 shows the positioning of the target on the Earth's surface according to the model of the Earth's ellipsoid and the errors that arise due to the difference in the geodetic heights of the point from which the drone took off and the point where the detected person is located. In the given scenario, the SAR operation would be carried out at position P' instead of at position P where the person is actually located.

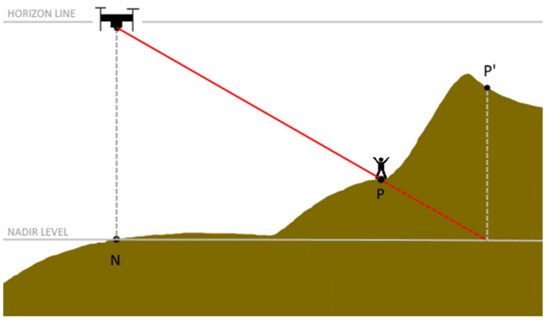

Figure 1: Schematic diagram of target geolocating error using the Earth ellipsoid model in areas with uneven terrain.

In the case of geographically complex terrains, data that rely on the Digital Elevation Model (DEM) (El Habchi et al., 2020), (Huang et al., 2020) can be used. DEM includes a database of the height of any location on Earth, expressed in relation to sea level. In (Paulin et al., 2024) a methodology for precise geolocation using DEM and the RayCast method was introduced and it was shown that the use of DEM significantly increases the accuracy of person positioning on complex terrain.

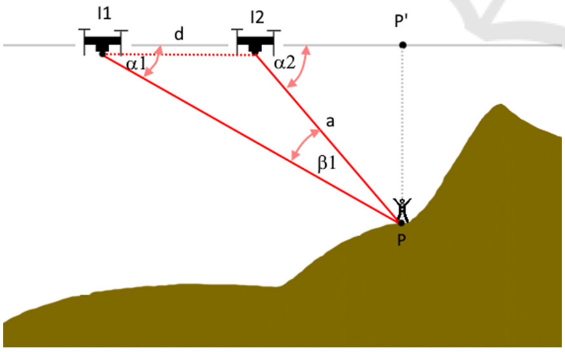

Figure 2: Two-point intersection positioning model.

Another approach focused in reducing the elevation error includes two-point shooting on known GPS positions (I1 and I2 on Fig. 2) at a single target and a direction vector that usually depends on angle sensor of drone camera (Qu et al., 2013), (Xu et al., 2020). This algorithm can only be used for geolocation of stationary targets because its accuracy is significantly degraded when the target moves. The solution is the approach in (Bai et al., 2017), which uses two drones at positions I1 and I2, for simultaneous recording of the same target and determination of the cross-section and the position of the target. However, this algorithm is not applicable for the case of SAR due to the additional cost of the drone that should record the same search area and due to the safety issue where the simultaneous use of the same airspace by multiple drones is avoided to reduce the risk of collision.

3 PERSON DETECTION AND GEOLOCATION IN SAR MISSION

3.1 YOLOv8 for Person Detection

The YOLOv8 is engineered with a focus on improving performance of real-time detection of objects of various sizes while reducing inference time and computing requirements (Ultralytics, n.d.-c) which makes it potentially interesting for use in SAR missions that generally have small objects of interest and limited resources.

The YOLOv8 is presented in five distinct scaled versions with different number of free parameters: YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x. The YOLOv8n has the simplest architecture with 3 million parameters, while YOLOv8x, has 68 million parameters and shows the best performance within the shortest time (Table 1.).

Table 1: Comparison of five YOLOv8 models, trained and evaluated on the COCO test-dev 2017 dataset with 640 px input, according to the report from (Ultralytics, n.d.-b).

Version of YOLO	mAP 50-95	Speed CPU ONNIX (ms)	Speed A100 Tensor RT (ms)	params (M)
YOLOv8n	37.3	80.4	0.99	3.2
YOLOv8s	44.9	128.4	1.20	11.2
YOLOv8m	50.2	234.7	1.83	25.9
YOLOv8l	52.9	375.2	2.39	43.7
YOLOv8x	53.9	479.1	3.53	68.2

We have fine-tuned all five versions of the YOLOv8 model on the SARD dataset adapted for object detection in SAR with two changes to the original architecture: the input to the network was changed to dimensions of 640 for images of 640x360 pixels, and the output, to one class (a person).

3.2 Geolocation Estimation

In SAR missions, it is very often the case that missing persons are motionless because they are injured and/or exhausted. Therefore, we propose a geolocation intersection measurement algorithm for locating missing person, that relies on the analysis of multiple shots taken by a single drone and uses terrain configuration data to reduce geolocation error. The algorithm starts to be used after a person is detected in an image, and then an intersection is determined with each subsequent image in which there is also a detected person. In Figure 2, label d is the distance between two drone positions from which the images were captured. Angles α1 and α2 are determined in the same manner as in (Sambolek & Ivašić-Kos, n.d.). By applying the same rule, we calculate the length of side I1P, which is the distance from the drone to the person (point P) when the first image was taken, and the length of side I2P (length a in Figure 2, equation 1), represents the distance from the location where the second image was taken. Then, from the triangle I2PP', we determine the length of side I2P' (Eq. 2), based on which we calculate the GPS coordinates of point P, considering known GPS coordinates of the drone's position and the azimuth toward point P.

$$\frac{a}{\sin\alpha 1} = \frac{d}{\sin\beta 1} \tag{1}$$

$$\frac{a}{\sin\alpha 1} = \frac{d}{\sin\beta 1} \tag{2}$$

Geolocation results is the distance in meters between two points at Earth according to the current standard WGS 84 that is reference system used by the GPS and identifies an Earth-centered, Earth-fixed coordinate system with absolute accuracy of 1-2 meters. The mean error (Eq. 4) indicates the average value of all distances ΔPi (Eq. 3) calculated between predicted geolocation of detected points and the GT point, P_{GT_i} for each image in the dataset.

$$\Delta P_i = Geodesic.WGS84.Inverse(P_i, P_{GT_i}) \tag{3}$$

$$Mean\ Error = \frac{\sum_{i=1}^{n} \Delta P_i}{n} \tag{4}$$

4 EXPERIMENTS

4.1 Datasets

In our study, we used two datasets, SARD and SAR-DAG_overflight. The SARD dataset was used for training the YOLOv8 model for person detection, while the SAR-DAG_overflight dataset was prepared for the validation of the geolocation algorithm of detected persons.

4.1.1 SARD - Dataset for Training Detector

The SARD dataset was designed with a particular focus on detecting missing or injured persons captured by drones in non-urban terrains. The data was recorded by a DJI Phantom 4 Advanced drone in continental Croatia and includes 1,981 images with a total of 6,532 people. Examples of images from the SARD set are shown in Figure 3.

Figure 3: Examples of detections on images from the SARD dataset with an enlarged image to better emphasize the person in the image that needs to be detected.

The images from the SARD set are of 640 x 360 resolution and are evenly distributed in a ratio of 60:40 into a training set and a validation set based on various factors such as background, lighting, person pose, and camera angle. The training set contains 1,189 images with 3,921 tagged persons, while the validation set contains 792 images with 2,611 tagged persons (Sambolek & Ivasic-Kos, 2021).

In this experiment, we removed from the training set all images that contained a frame with a person with an area of less than 102 pixels, which significantly saved the amount of computer time during training without negatively affecting the performance of the model. After this intervention, the training set contains 817 images with 2017 people, of which 1779 are small objects (area < 322 pixels) and 238 medium objects (area between 322 and 962 pixels), while there are no large objects (area > 962 pixels).

4.1.2 SAR-DAG_Overflight - Datasets for Evaluating Geolocation Method

To test the geolocation algorithm, we created a set of images taken at two locations, a meadow, and a vineyard. The images were captured by a Phantom 4

Advance drone, equipped with a camera with a field of view of 84° that flew at a height of 30 meters and captured images at regular time intervals as is usual in SAR missions. The images have a resolution of 5472 x 3648 pixels, and an example is given in Figure 4. The set contains 40 marked persons. From the metadata of the images taken at the position where the drone took off and at the position when the drone is vertically above the person, GPS position data is taken to obtain the starting point and the actual position of the person on the ground.

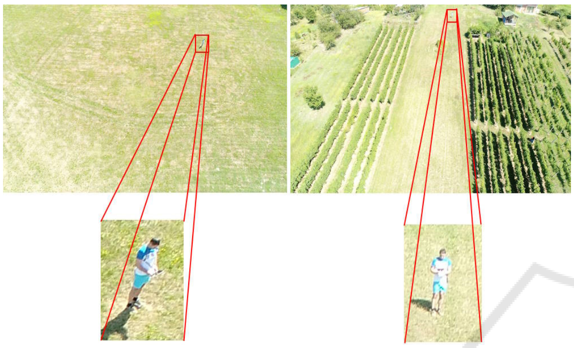

Figure 4: Examples of SAR-DAG_overflight images with zooming in on a part of the image where the person is.

4.2 Evaluation Metric

In the experiment, we use several standard metrics to evaluate detector performance and metrics that we have purpose-developed for detection and geolocation in SAR missions as explained below.

Intersection over Union (IoU) is a traditional metric for evaluating the performance of an object detector calculated as the ratio of the intersection and union of the detected bounding box and the ground true bounding box. The equation is as follows:

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union} \tag{5}$$

Higher IoU values indicate better overlap between detection and the real data.

Recall (R) and Precision (P) are calculated as:

$$P = \frac{TP}{TP + FP}, \qquad R = \frac{TP}{TP + FN} \tag{6}$$

where TP is positive detection that are true, FP is false positives, and FN is false negative detection.

Mean average precision (mAP) is a common evaluation metric in object detection. In the experiment, we use mAP 50, the average precision at IoU greater than or equal to 0.5 and mAP 50-95 the average precision in the range of IoU from 0.5 to 0.95, with intervals of 0.05.

For SAR operations, it is important that the detector is optimized to have as few false positive (FP) detections as possible, because they consume human resources and time. Therefore, the performance of the detector is also evaluated using the ROpti (Recall Optimal) metric, which penalizes false positive detections (Sambolek & Ivasic-Kos, 2021). ROpti is calculated as the ratio of the difference between true positive (TP) and false positive (FP) detections and the total number of detections (TP+FN):

$$ROpti = \frac{TP - FP}{TP + FN} \tag{7}$$

The experiments also evaluate the accuracy of geolocating a person using the proposed algorithm (Section 3.2).

4.3 Experimental Results

4.3.1 YOLOv8 Person Detection

We conducted the experiments using all five versions of the YOLOv8 models modified to detect a person class and implemented in PyTorch using Python version 3.9.16.

First, on the SARD validation set we tested original YOLOv8 models trained on the COCO dataset, and the obtained results are shown in Table 2. The confidence threshold was set to 0.25 and the IoU threshold to 0.5.

The YOLOv8x model achieved the best result of all YOLOv8 versions on the SARD validation set, namely mAP@0,5 of 74.6%, recall of 49.2%, and mAP@ 0.5:0.95 of 35%, which is significantly worse than when tested on the COCO set. Although it is a simplified detection task with only one class (person), all YOLOv8 models show the same performance degradation with many false detections (low ROpti). Considering that the SARD set was recorded from a completely different perspective (bird's eye view) and with many small objects for which the models were not trained, it was necessary to fine-tune them to SARD datasets so that they can be used in SAR missions.

We trained all version of YOLOv8 models for 500 epochs using Tesla T4 GPUs on the Google Collaboratory platform while the hyperparameters remained unchanged. We used SGD optimizer, and the weight decay set to 5 x 10^{-4}, while the initial learning rate was set to 10^{-3}. Input image size was 640 and batch size set to 16.

Detection performances on SARD validation dataset were evaluated using standard metrics of

Precision, Recall, mAP@0.5, and mAP@0.5:0.95, and customized ROpti measure (Sambolek & Ivasic-Kos, 2021). After fine-tuning on the SARD data set all models show a significant improvement in detection (Table 2.). The best results were achieved by YOLOv8x with mAP@0.5 91.3% and mAP@0.5:0.95 68.8%, which makes it the most suitable for offline analysis of materials recorded during drone flight because the accuracy is in that case the most important.

The YOLOv8n model has the significantly fastest detection of only 4.6 ms per image and achieves mAP@0.5 only 4.5% lower than the best results. The same is true for the YOLOv8s model, which achieves the second-best inference time with almost the same mAP@0.5 performance as YOLOv8x. This makes it most suitable for use during a SAR operation when, in addition to detection accuracy, it is important for the model to inference quickly, in real time, and to be used on a drone without the need for large computing resources.

4.3.2 Person Geolocation

We have conducted a comparison of existing geolocation methods using a simplified ellipsoidal model of the Earth, an algorithm using DEM (Digital Elevation Model) and an intersection measurement algorithm. The results of the first two measurements were taken from the paper (Sambolek & Ivašić-Kos, n.d.). Table 3 shows the results of the distance estimation between the calculated GPS location of a person using the mentioned three algorithms and the exact GPS location where the person was located. The algorithms were tested on five different data sets, two of which were recorded in a meadow (flat terrain), while three were recorded in a vineyard (sloping terrain). In data sets recorded in the meadow, no major deviation was observed for intersection algorithms that consider changes in the terrain configuration (e.g., a mean error of 4.5 m for PhantomLP1), however, on terrains with different slopes, the intersection measurement algorithm shows significantly better results than other algorithms.

The best result was achieved in the first set recorded in the vineyard (PhantomVP1), with an average error of 4.8 meters. In the case of the Earth ellipsoid model and the DEM model, accuracy was checked for each image in the dataset.

If a person is detected in one image or is in motion during the search, it is recommended to use the DEM model to determine the geolocation. When detecting a stationary person in multiple images, it is suggested to use the intersection measurement algorithm, which achieves the best results.

Table 2: Performance of five versions of the YOLOv8 model on the SARD test dataset. The first five rows correspond to models trained on the COCO dataset and the last five to models that are fine-tuned on the SARD dataset, with the best results highlighted in bold.

Version of YOLO and training dataset	Precision (%)	Recall (%)	mAP @0,5 (%)	mAP @ 0.5:0.95 (%)	ROpti	Speed per image [ms]
YOLOv8n @COCO	61	26	35.9	16.5	0.09	**4,8**
YOLOv8s @COCO	66	37	47.5	23.8	0.18	8,5
YOLOv8m @COCO	74	46	59.6	32	0.29	17.5
YOLOv8l@ COCO	**75**	47	60.7	34.5	0.31	34.5
YOLOv8x @COCO	**75**	**49**	**62.0**	**35.3**	**0.32**	46.6
YOLOv8n @SARD	93	78	86.8	54.9	0.71	**4.6**
YOLOv8s @SARD	94	81	90.3	60.6	0.76	8.0
YOLOv8m @SARD	93	83	90.6	62.1	0.77	17.3
YOLOv8l@ SARD	94	83	90.8	60.8	0.78	34.4
YOLOv8x @SARD	**95**	**83**	**91.3**	**63.8**	**0.79**	46.5

Table 3: Coordinates calculation of person standing on a known location.

Dataset	No. of Images	Earth ellipsoid model (Sambolek & Ivašić-Kos, n.d.)			DEM (Sambolek & Ivašić-Kos, n.d.)			Intersection measurement algorithm		
		MeanError	MaxError	MinError	MeanError	MaxError	MinError	MeanError	MaxError	MinError
PhantomLP1	10	8.963	10.539	7.87				13.446	14.377	12.713
PhantomLP2	10	8.704	11.595	6.212				8.439	8.832	7.592
PhantomVP1	4	18.374	29.262	8.412	10.935	15.833	5.630	**4.794**	5.451	4.004
PhantomVP2	7	50.488	73.028	14.427	23.604	34.681	7.327	10.534	11.139	10.351
PhantomVP3	9	51.312	98.203	22.815	29.911	66.887	14.762	12.388	14.465	9.725

5 CONCLUSIONS

In this paper, we have demonstrated that the YOLOv8 models can be successfully fine-tuned on UAV images for person detection in real-world environments. Our experiment was conducted on the publicly available SARD dataset.

Furthermore, we built a set of SAR-DAG_overflight for testing the geolocation of a person and tested three geolocation algorithms on it: the Earth's ellipsoid model, the DEM model, and the modified cross-section measurement algorithm that we proposed in the paper.

We believe that the fine-tuned YOLOv8@SARD models that we fine-tuned at the SARD dataset and the proposed person geolocation algorithms along with the given recommendations can be greatly utilized in SAR operations as they can help in the detection of persons in drone images, and thus contribute to providing more precise information for coordinating the operation and reducing search time.

In future work, we plan to further investigate the model's robustness to weather conditions, night shooting, and camera motion blur, as well as conduct experiments with multiple datasets to increase the robustness and generalizability of our model.

ACKNOWLEDGMENTS

This research was partially supported by HORIZON EUROPE Widening INNO2MARE project (grant agreement ID: 101087348).

REFERENCES

Andriluka, M., Schnitzspan, P., Meyer, J., Kohlbrecher, S., Petersen, K., Von Stryk, O., Roth, S., & Schiele, B. (2010). Vision based victim detection from unmanned aerial vehicles. *IEEE/RSJ 2010 International Conference on Intelligent Robots and Systems, IROS 2010 - Conference Proceedings*. https://doi.org/10.1109/IROS.2010.5649223

Bai, G., Liu, J., Song, Y., & Zuo, Y. (2017). Two-UAV intersection localization system based on the airborne optoelectronic platform. *Sensors (Switzerland)*, *17*(1). https://doi.org/10.3390/s17010098

Bejiga, M. B., Zeggada, A., Nouffidj, A., & Melgani, F. (2017). A convolutional neural network approach for assisting avalanche search and rescue operations with UAV imagery. *Remote Sensing*, *9*(2). https://doi.org/10.3390/rs9020100

Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020). *Yolov4: Optimal speed and accuracy of object detection*.

Chien-Yao Wang, Alexey Bochkovskiy, H.-Y. M. L. (2023). YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 7464–7475.

Doherty, P., & Rudol, P. (2007). A UAV search and rescue scenario with human body detection and geolocalization. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *4830 LNAI*. https://doi.org/10.1007/978-3-540-76928-6_1

El Habchi, A., Moumen, Y., Zerrouk, I., Khiati, W., Berrich, J., & Bouchentouf, T. (2020). CGA: A New Approach to Estimate the Geolocation of a Ground Target from Drone Aerial Imagery. In *4th International Conference on Intelligent Computing in Data Sciences, ICDS 2020*. https://doi.org/10.1109/ICDS50568.2020.9268749

Geraldes, R., Goncalves, A., Lai, T., Villerabel, M., Deng, W., Salta, A., Nakayama, K., Matsuo, Y., & Prendinger, H. (2019). UAV-based situational awareness system using deep learning. *IEEE Access*, *7*. https://doi.org/10.1109/ACCESS.2019.2938249

Huang, C., Zhang, H., & Zhao, J. (2020). High-efficiency determination of coastline by combination of tidal level and coastal zone DEM from UAV tilt photogrammetry. *Remote Sensing*, *12*(14). https://doi.org/10.3390/rs12142189

Leira, F. S., Trnka, K., Fossen, T. I., & Johansen, T. A. (2015). A ligth-weight thermal camera payload with georeferencing capabilities for small fixed-wing UAVs. *2015 International Conference on Unmanned Aircraft Systems, ICUAS 2015*. https://doi.org/10.1109/ICUAS.2015.7152327

Li Chuyi, Li Lulu, Jiang Hongliang, Weng Kaiheng, Geng Yifei, Li Liang, Zaidan Ke, Qingyuan Li, Meng Cheng, Weiqiang Nie, Yiduo Li, Bo Zhang, Yufei Liang, Linyuan Zhou, Xiaoming Xu, Xiangxiang Chu, Xiaoming Wei, X. W. (2022). *YOLOv6: A single-stage object detection framework for industrial applications.*

Paulin, G., Sambolek, S., & Ivasic-Kos, M. (2024). Application of raycast method for person geolocalization and distance determination using UAV images in Real-World land search and rescue scenarios. *Expert Systems with Applications*, *237*. https://doi.org/https://doi.org/10.1016/j.eswa.2023.121495

Qu, Y., Wu, J., & Zhang, Y. (2013). Cooperative localization based on the azimuth angles among multiple UAVs. *2013 International Conference on Unmanned Aircraft Systems, ICUAS 2013 - Conference Proceedings*. https://doi.org/10.1109/ICUAS.2013.6564765

RangeKing. (n.d.). *YOLO v8 architecture*. https://github.com/ultralytics/ultralytics/issues/189

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2016-December*. https://doi.org/10.1109/CVPR.2016.91

Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. *Tech Report*.

Redmon, J., & Farhadi, A. (2017). YOLO9000: Better, faster, stronger. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, *2017-January*. https://doi.org/10.1109/CVPR.2017.690

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(6). https://doi.org/10.1109/TPAMI.2016.2577031

Sambolek, S., & Ivasic-Kos, M. (2021). Automatic person detection in search and rescue operations using deep CNN detectors. *IEEE Access*, *9*, 37905–37922. https://doi.org/10.1109/ACCESS.2021.3063681

Sambolek, S., & Ivašić-Kos, M. (n.d.). *Determining the Geolocation of a Person Detected in an Image Taken with a Drone*.

Shakhatreh, H., Sawalmeh, A. H., Al-Fuqaha, A., Dou, Z., Almaita, E., Khalil, I., Othman, N. S., Khreishah, A., & Guizani, M. (2019). Unmanned Aerial Vehicles (UAVs): A Survey on Civil Applications and Key Research Challenges. In *IEEE Access* (Vol. 7). https://doi.org/10.1109/ACCESS.2019.2909530

Sun, J., Li, B., Jiang, Y., & Wen, C. Y. (2016). A camera-based target detection and positioning UAV system for search and rescue (SAR) purposes. *Sensors (Switzerland)*, *16*(11). https://doi.org/10.3390/s16111778

Ultralytics. (n.d.-a). *Yolov5 GitHub*. Retrieved September 15, 2023, from https://github.com/ultralytics/yolov5

Ultralytics. (n.d.-b). *YOLOv8 Doc*. https://docs.ultralytics.com/tasks/detect/

Ultralytics. (n.d.-c). *YOLOv8 GitHub*. Retrieved September 15, 2023, from https://github.com/ultralytics/ultralytics

Wang, X., Liu, J., & Zhou, Q. (2017). Real-time multi-target localization from unmanned aerial vehicles. *Sensors (Switzerland)*, *17*(1). https://doi.org/10.3390/s17010033

Xu, C., Yin, C., Han, W., & Wang, D. (2020). Two-UAV trajectory planning for cooperative target locating based on airborne visual tracking platform. *Electronics Letters*, *56*(6). https://doi.org/10.1049/el.2019.3577

Zhao, X., Pu, F., Wang, Z., Chen, H., & Xu, Z. (2019). Detection, tracking, and geolocation of moving vehicle from UAV using monocular camera. *IEEE Access*, *7*. https://doi.org/10.1109/ACCESS.2019.2929760

Zou, Z., Chen, K., Shi, Z., Guo, Y., & Ye, J. (2023). Object Detection in 20 Years: A Survey. *Proceedings of the IEEE*, *111*(3). https://doi.org/10.1109/JPROC.2023.3238524.