Human detection with YOLO for last-mile delivery applications using UAVs

Débora Paula Simões [1,3], Henrique Cândido de Oliveira [2], Rafael Lino dos Santos [1]

[1] Graduate Program in Civil Engineering, School of Civil Engineering, Architecture and Urban Design, University of Campinas – Unicamp, Campinas - SP, Brazil - (d263621, r228749)@dac.unicamp.br

[2] Department of Infrastructure and Environment, School of Civil Engineering, Architecture and Urban Design, University of Campinas – Unicamp, Campinas - SP, Brazil - hcandido@unicamp.br

[3] Department of Surveying and Cartography, Federal Institute of Education, Science and Technology of Southern Minas Gerais – IFSULDEMINAS, Inconfidentes - MG, Brazil – debora.simoes@ifsuldeminas.edu.br

Keywords: Deep learning, drone, photogrammetry, logistics, path planning.

Abstract

Low-cost alternative solutions have advanced last-mile delivery, with Unmanned Aerial Vehicles (UAVs) emerging as a promising option for logistics tasks. However, as UAV operations increasingly occur in densely crowded urban areas, safety concerns - especially for people nearby - have intensified. To ensure safe deliveries, real-time UAV path planning is essential for avoiding no-fly zones defined around individuals detected along the route. This study addresses this challenge by evaluating human detection confidence in UAV imagery using the YOLOv7 model and a custom dataset. It also estimates individuals' positions through the Monoplotting technique. The customized YOLOv7 model achieved an average precision of 53.8% and an inference time of 9.9 ms, supporting real-time deployment. However, challenges in UAV-based human detection significantly influenced detection confidence. In one scenario, confidence exceeded 85% for individuals identified at a flight height of 30 m, while in another, the highest confidence reached 60% for imagery captured at 20 m. Despite numerous false negatives, the individual closest to the UAV was consistently detected, underscoring the applicability of the method for real-time path replanning. Regarding coordinate estimation of detected individuals, the Inertial Measurement Unit (IMU) system exerted the greatest influence on the accuracy of 3D positions obtained through Monoplotting. The phototriangulation process, which provided the sensor orientation parameters, directly impacted the results. To reduce discrepancies between estimated and actual coordinates, mathematical correction strategies may be applied; however, these require further investigation. The integration of Simultaneous Localization and Mapping (SLAM) techniques offers a promising direction for refining UAV rotation angles.

1. Introduction

With the rapid pace of technological advancement, low-cost alternative solutions have improved last-mile delivery. The use of Unmanned Aerial Vehicles (UAVs) is promising in logistics operations, with a projected growth of 53.8% by 2030 (Raja et al., 2023). In this context, UAVs are considered one of the essential technologies for implementing cost-effective and efficient infrastructures for smart cities (Othman and Aydin, 2023).

Conversely, as these operations are often conducted in densely populated urban areas, concerns about airspace safety and, particularly, the safety of people near the operation sites are significant. To mitigate risks, UAV flights over people are prohibited in many countries, and a safe distance between UAVs and the population is required (Australian Government, 2023; DECEA, 2023). Therefore, to ensure safe operations, the positions of individuals detected along the UAV's route must be estimated so that the UAV's trajectory avoids the no-fly zones defined around people.

With advances in Artificial Intelligence in recent years, several deep learning models have been developed, enabling human detection in UAV imagery. Among the most prominent are the YOLO (You Only Look Once) family of models – adopted by Bachir & Memon (2022) and Serghei et al. (2023), for instance, for person detection in UAV imagery. However, human detection in UAV images remains a challenging task due to the target's particularities and the conditions of the aerial images obtained.

Small objects, such as humans, in complex backgrounds, unstable environments, and under varying lighting conditions (Zhang et al., 2020), captured by cameras at high altitudes and with different viewing angles (Golcarenarenji et al., 2021), are difficult to detect. Moreover, the diversity of human poses (Golcarenarenji et al., 2021), the large volume of UAV image data, and issues related to camera motion, brightness, and blur, which affect image quality, further complicate human detection (Agarwal et al., 2021; Dousai and Lonearic, 2022). Detecting individuals in crowded areas, such as urban centers where last-mile delivery operations with UAVs are typically conducted, is an especially challenging task (Symeonidis et al., 2022).

Another issue specific to this task is the scarcity of publicly available labeled datasets (Guettala et al., 2022), which are crucial for training deep learning models capable of object detection (Dousai and Lonearic, 2022). Although there are already specific UAV datasets for human detection, such as VisDrone (Zhu et al., 2022) and Manipal-UAV (Akshatha et al., 2023), finding a suitable dataset for a specific application, on a large scale, and with accurate annotations remains a significant challenge (Sun et al., 2022).

In light of these bottlenecks, the present study aims to evaluate the accuracy and speed of human detection in UAV images using deep learning, adopting a custom dataset to train the YOLOv7 model. Considering that human detection is essential for planning safe UAV routes for last-mile delivery, this study also examines the accuracy of the Monoplotting photogrammetric process when

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-3-W4-2025-349-2026

estimate the georeferenced three-dimensional coordinates of the individual closest to the UAV detected by the deep learning model.

This article is structured as follows: Section 2 provides details about the deep learning model and the dataset created for human detection in UAV images in this section. Additionally, the method employed to estimate the detected person's position is described. Section 3 presents the experiments conducted and discusses the achieved accuracy. Section 4 addresses the challenges identified in this research and highlights opportunities for future work on safe UAV route planning. Finally, Section 5 presents the final considerations regarding the proposed method.

2. Human detection and position estimation

Figure 1 provides a general overview of the proposed approach to ensure safe path planning for UAVs employed in last-mile delivery operations.

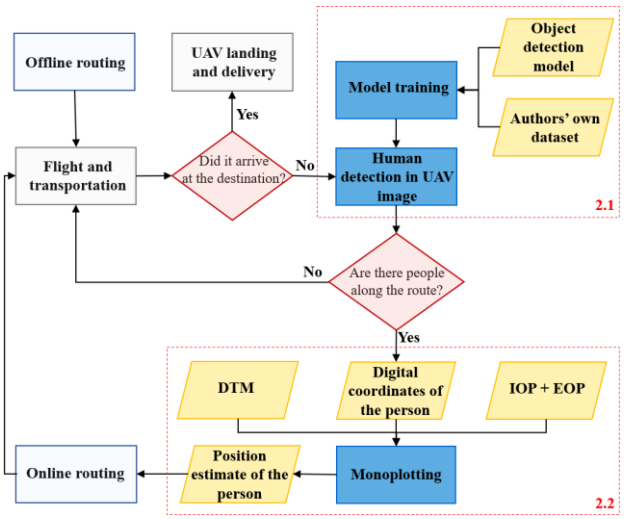

Figure 1. Overview of the proposed method for safe UAV path planning.

The process begins with the definition of a preliminary UAV route. Details of the offline routing procedure are provided by Simões et al. (2022). While this offline route can avoid no-fly zones around previously mapped buildings, the UAV may still fly over non-consenting individuals, thereby posing a safety risk. To address this issue, the YOLOv7 model (Wang et al., 2022) is employed for real-time human detection, as outlined in Section 2.1. If a person is detected within an unsafe distance from the UAV, the route is dynamically replanned along the trajectory, according to airspace access regulations. To generate the updated (online) route, it is necessary to estimate the georeferenced three-dimensional coordinates of the detected individual. This is accomplished using the Monoplotting photogrammetric technique (Fluehler et al., 2005), described in Section 2.2.

2.1 Human detection in a UAV image

For human detection in UAV images, the YOLOv7 detection algorithm (Wang et al., 2022) was adopted due to its real-time performance and high accuracy (Kamath and Renuka, 2023). This model is designed for real-time operations on mobile devices using GPUs (Graphics Processing Units) (Serghei et al., 2023; Wang et al., 2022). As a result, it can be optimized for embedded systems (Serghei et al., 2022), making it particularly suitable for last-mile delivery operations using drones. Beyond these technical advantages and features, the decision to use this deep learning model is also supported by the fact that YOLO-based algorithms currently represent the state of the art in real-time object detection (Song et al., 2024). In particular, YOLOv7 has been adopted by several researchers and has produced satisfactory results for human detection in UAV images (Serghei et al., 2023, 2022; Song et al., 2024).

However, the YOLOv7 model is trained on the COCO dataset (Lin et al., 2014), which includes 91 object classes based on ground-level images. As a result, the pre-trained YOLOv7 model is not directly suitable for the present study. To address this, transfer learning was employed (Zhang et al., 2023). Using the pre-trained weights, YOLOv7 was retrained on a custom dataset (Section 2.1.1) through a technique known as fine-tuning (Zhang et al., 2023). The training algorithms and parameters used for the pre-trained model were obtained from the official YOLOv7 repository on GitHub (Wong, 2022). Table 1 presents the main hyperparameters used in the customized training process.

Image size (pixels)	Number of epochs	Batch size	Workers
640 x 640	300	16	4

Table 1. Customized training hyperparameters - YOLOv7.

2.1.1 Unicamp-UAV dataset

Considering the significant impact of dataset quality on object detection accuracy, a custom dataset representative of the study area was developed for the customized training of the YOLOv7 model. This process included data collection and annotation, as outlined by Vijayakumar and Vairavasundaram (2024). Image acquisition was carried out using a DJI Phantom 4 UAV, with three videos recorded in the study area during daylight hours and under favorable weather conditions. Additionally, to ensure dataset heterogeneity and representativeness, various camera orientations were used, and individuals were captured in diverse poses.

After extracting the frames (color images with a resolution of 3840×2160 pixels – see Figure 2), data preprocessing and cleaning were performed. These steps are essential for reducing dataset complexity, preventing the model from learning incorrect features, and enhancing overall accuracy (Gupta and Verma, 2022). The final dataset consists of 6,000 positive images (with people) and 500 negative images (background only). Manual annotation of individuals in the UAV imagery was performed using the LabelImg tool (Sell, 2024), resulting in a total of 58,555 labeled instances.

Figure 2. Examples of frames extracted from videos recorded at the Barão Geraldo Campus, Unicamp, Brazil.

The 6,500 images comprising the "Unicamp-UAV" dataset were split into training (80%), which includes the validation dataset, and testing (20%), following the same proportions adopted by several authors, such as Sinha and Kumar (2023). The training

https://doi.org/10.5194/isprs-annals-X-3-W4-2025-349-2026

results were evaluated using the metrics precision (P), recall (R), average precision (AP), and speed measured in frames per second (FPS) (Zhang et al., 2020).

2.2 Real-time position estimation of a person

The UAV's flight path during last-mile delivery operations is established using waypoints defined by georeferenced three-dimensional coordinates. To enable route replanning when a person is detected during flight, it is essential to determine the individual's position within the same coordinate system. For this purpose, the Monoplotting photogrammetric procedure was employed (Fluehler et al., 2005). In this process, the initial coordinates of a point (X_i, Y_i) are calculated using the Inverse Collinearity Equations (Mikhail et al., 2001), considering the maximum elevation of the digital terrain model (DTM) within the area of interest ($Z_i = Z\_max$). A new elevation value (Z_{i+1}) is then defined by interpolating the initial coordinates with the DTM. By applying the Inverse Collinearity Equations again, a new position (X_{i+1}, Y_{i+1}) is calculated (Simões et al., 2023). This is an iterative process, continuing until a threshold is met (Fluehler et al., 2005), as illustrated in Figure 3(c).

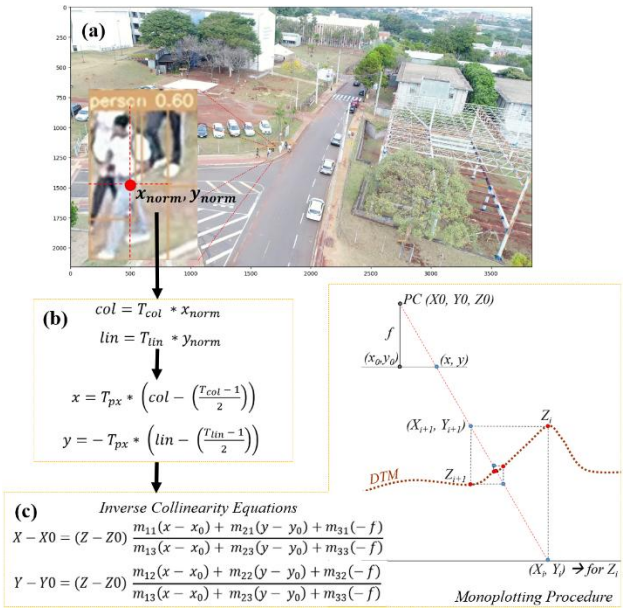

Figure 3. Position Estimation: (a) output of human detection using YOLOv7; (b) conversion to photogrammetric coordinates; (c) Monoplotting procedure.

The DTM used in this process is derived from a 3D point cloud of the area of interest, containing exclusively ground points. The LAStools software (Rapidlasso, 2024) was employed to filter the point cloud representing the Digital Surface Model (DSM), which was obtained using an Airborne Laser Scanning System (ALS).

To compute a person's position (X, Y) using the Inverse Collinearity Equations, the Interior Orientation Parameters (IOP) of the sensor used to capture the images are required. These parameters include the displacement of the principal point in the image (x_0, y_0) and the focal length (f), which are available in the UAV manufacturer's manual. Additionally, the Exterior Orientation Parameters (EOP) of the image in which the person was detected along the UAV's trajectory are necessary. These include the coordinates of the perspective center (PC) (X_0, Y_0, Z_0) and the sensor's attitude angles (κ, ϕ, ω) at the moment each image is acquired.

As a result of the human detection step using the YOLOv7 model, the normalized coordinates (x_{norm}, y_{norm}) of the center of the bounding box surrounding the detected person in the UAV image are obtained (see Figure 3(a)). These coordinates are then converted to digital system coordinates (column, line) and then converted into photogrammetric coordinates (x, y) (see Figure 3(b)). From these photogrammetric coordinates, the Inverse Collinearity Equations can be applied.

3. Experiments and discussion

3.1 Customized training of the YOLOv7 model

The training of the YOLOv7 model using the Unicamp-UAV dataset resulted in the following evaluation metrics on the test dataset: P = 91.3%; R = 54.4%; AP@.5 = 53.8%; and an inference speed of 9.9 ms (equivalent to 75.2 FPS). Although the training results are not optimal, particularly due to the Recall value, which indicates a high rate of false negatives (i.e., missed detections), the achieved P, R, and AP@.5 metrics surpass those reported in other studies involving YOLOv7 model training, such as the study presented by Liu et al. (2022). Furthermore, the FPS rate exceeds that reported in other studies focused on human detection in UAV images, such as the study by Liu and Szirányi (2021).

The inference of images obtained using the DJI Mavic 3M UAV had an average duration of 135.48 ms per image, utilizing a NVIDIA GeForce RTX 3060 6 GB GPU. Therefore, in terms of inference speed, the customized YOLOv7 model is capable of performing real-time detections and, therefore, it is well-suited for UAV path replanning in last-mile delivery operations.

3.2 Analysis of human detection confidence in UAV images

Using the weights trained on the Unicamp-UAV dataset, predictions were performed with the customized YOLOv7 model on images captured by a DJI Mavic 3M UAV at different flight heights and across two distinct scenarios. The detection confidence of the person closest to the UAV was assessed, as this individual's position is considered for real-time route replanning.

In the first experiment, five images were acquired at flight heights of 10, 20, 30, 40, and 50 meters. Figure 4 illustrates the detection of the individuals closest (Person I) and farthest (Person II) from the UAV in the image captured at a height of 30 meters, while Figure 5 presents the detection confidence for Person I across all images.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-3-W4-2025-349-2026 | © Author(s) 2026. CC BY 4.0 License.

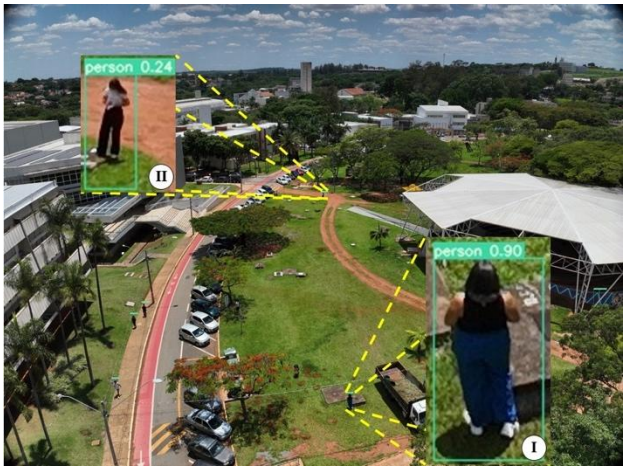

Figure 4. Human detection in a UAV image with YOLOv7 – Experiment I.

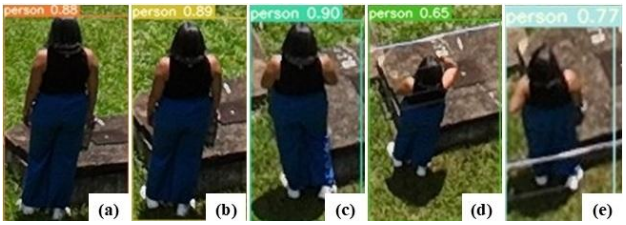

Figure 5. Detection confidence and analysis of Person I in UAV images at different flight heights: (a) 10 m; (b) 20 m; (c) 30 m; (d) 40 m; (e) 50 m.

The lowest detection confidence was observed in the image captured at a flight altitude of 40 meters. However, it was expected that the lowest confidence would occur in the image captured at 50 meters, given that higher altitudes typically reduce the size of the person in the image, making detection more challenging. It is presumed that the lower detection confidence for Person I at 40 meters is due to the individual's pose (Figure 5). This observation reinforces the notion that human pose is one of the most challenging factors affecting the accuracy of person detection in UAV images.

It can be observed that images captured at heights of 10 m, 20 m, and 30 m show high and consistent detection confidence levels. Based on this observation, it is recommended that for flights up to 30 m, a minimum confidence threshold of 85% should be applied for real-time model predictions. Since the inference parameter "conf" can be adjusted in the model's prediction mode, setting an appropriate threshold helps ensure that objects detected with low confidence are ignored, thereby reducing the number of false positives (FP) (Ultralytics, 2025). This is the case for the image acquired at an altitude of 10 meters (Figure 6): bicycles (yellow bounding box) were incorrectly detected as persons (FP). However, since these detections have confidence scores below the 85% threshold, they were excluded.

Figure 6. False Positives (yellow) and False Negatives (red).

For images captured above 30 m, detection confidence significantly decreases. This occurs due to the reduced size of the object, in this case, the person, in images taken from higher flight altitudes. A similar effect is observed when individuals appear farther from the UAV's position within the same image. For example, Figure 4 also shows the detection of Person II, who is the most distant individual from the UAV in the image. In this case, at a flight altitude of 30 meters, the detection confidence is only 24%. However, since UAV path replanning for secure last-mile delivery takes into account only the person closest to the UAV, this limitation does not compromise the proposed method.

Figure 6 also shows several individuals who were not detected (red bounding boxes), i.e., False Negatives (FN). However, since the person closest to the UAV – the only one considered in our UAV path replanning method – is detected in all evaluated images (Figure 5), the observed FNs do not compromise the proposed application. In other scenarios and applications, this could represent a significant concern. In Search and Rescue (SAR) operations, for instance, FNs can result in loss of life. It is important to highlight that, for the image in question (Figure 6), the missed detections occurred due to the individuals' small size and varying human poses. In other situations, partially occluded individuals – such as those hidden by cars, trees, or buildings – also represent targets that are difficult to detect.

In addition, UAV images with more complex backgrounds, varying lighting conditions, and lower contrast – due to the season in which they were captured – as well as differing rotation angles, further complicate human detection. This is evident in the second experiment (Figure 7). Considering the horizontal distance between the UAV's ground projection and the detected individual (Person III) of approximately 30 meters, four images were acquired at flight heights of 10, 20, 30, and 40 meters. For this scenario, the detection confidence was lower than that observed in the previous experiment (Figure 5) for all images due to the characteristics of the UAV-captured images. Even the highest confidence level (Figure 7 – b) remained below the lowest detection confidence recorded in the previous experiment (Figure 5 – d).

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-3-W4-2025-349-2026 | © Author(s) 2026. CC BY 4.0 License.

Figure 7. Detection confidence of Person III in UAV images at different flight heights: (a) 10 m; (b) 20 m; (c) 30 m; (d) 40 m.

To support the observation that image characteristics hinder detection in the second scenario, an additional image was acquired at a flight height of 20 meters – the same height as the image with the highest detection confidence in Experiment II. The horizontal distance between the UAV's ground projection and Person III was approximately 70 meters (Figure 8). Although the individual appears smaller due to the distance between Person III and the UAV, the detection confidence surpassed that of the previous 20-meter image shown in Figure 7 (b). These results reinforce the challenge of generalizing the YOLOv7 model for human detection in UAV imagery, given the variability in individual poses and the differences in image quality across scenarios.

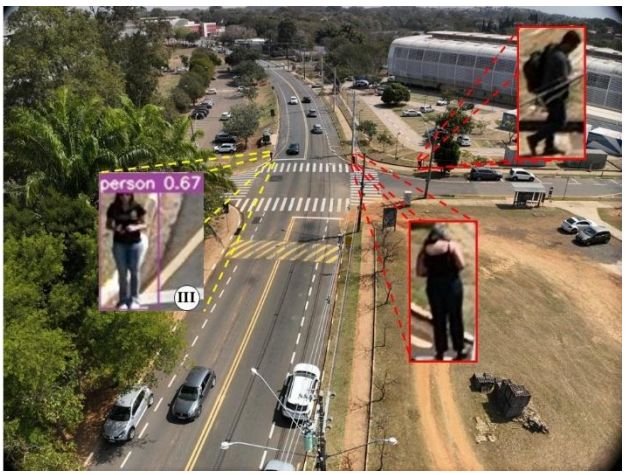

Figure 8. Human detection in UAV imagery at 20 m flight height – Experiment II.

Considering a detection confidence threshold of 60%, FNs are observed in Figure 8. As in the previous experiment, for the specific application under study – real-time replanning of safe UAV routes – these FNs would not compromise the method's effectiveness, since a person in close proximity to the undetected individuals was correctly identified. This would still trigger a deviation of the UAV from that area, due to the no-fly zone established around Person III. However, the results obtained in Experiment II reinforce the fact that, for other applications, further research is needed, particularly to properly adjust the "conf" parameter in the prediction process of the deep learning model.

3.3 Analysis of accuracy for human position estimation

Based on the YOLOv7 prediction results from the five images from Experiment I (Figure 5), the georeferenced 3D coordinates of Person I were calculated and compared with their actual coordinates, which were obtained from a true orthophotomosaic with a spatial resolution of 7 cm. The discrepancies are presented in Table 2.

Image	Flight altitude	ΔX (m)	ΔY (m)	ΔZ (m)	Planimetric Error (m)
1	10 m	5.193	3.988	-0.292	6.548
2	20 m	3.084	2.570	-0.208	4.014
3	30 m	2.382	2.493	-0.089	3.448
4	40 m	2.401	2.287	-0.154	3.316
5	50 m	2.828	3.046	-0.216	4.156
			Mean error (m)	-0.192	4.296

Table 2. Difference between calculated and real coordinates – Experiment I.

The errors observed in estimating Person I's coordinates using the Monoplotting procedure are directly proportional to the distance between the UAV's position at the moment the image was captured and the estimated position of the individual (Figure 9), which is consistent with the findings of Simões et al. (2023).

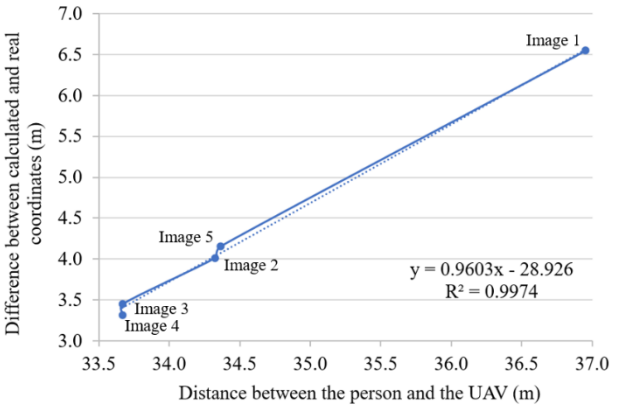

Figure 9. Simple linear regression – Experiment I.

The planimetric error is a result of the accuracy of the UAV's Inertial Measurement Unit (IMU) positioning system. The DJI Mavic 3M UAV is equipped with a Global Navigation Satellite System (GNSS) Real Time Kinematic (RTK) positioning system (DJI, 2025), which provides high precision in determining the coordinates of the perspective centers (X0, Y0, Z0) of the images used for human detection. These EOPs were extracted directly from the images' EXIF (Exchangeable Image File Format) metadata. However, the UAV's inertial system does not offer the same level of precision in estimating the camera's attitude – that is, the navigation angles (roll, pitch, and yaw).

Even when the Euler angles (κ, ϕ, ω) are obtained through the phototriangulation process, the planimetric errors in estimating the position of Person I using Monoplotting remain significant. This is because, in UAVs equipped with high-precision GNSS positioning systems, the IMU contributes to approximately 99%

of the accuracy in projecting planimetric coordinates (Simões et al., 2023). Therefore, when estimating coordinates via Monoplotting, the EOPs related to the sensor's attitude (κ, ϕ, ω) are highly sensitive parameters; even small variations in their values can result in high magnitude errors.

It is important to highlight that when calculating the coordinates of Person II (Figure 4), who is located 129.593 m from the UAV, the resulting planimetric error is 30.546 m. This value does not align with the line equation derived from the values related to Person I (Figure 9).

Considering the detection results for the four images from Experiment II (Figure 7), the 3D georeferenced coordinates of Person III were computed. Table 3 presents the differences between the position of Person III as determined through monoplotting and that obtained using the true orthophotomosaic.

Image	Flight altitude	ΔX (m)	ΔY (m)	ΔZ (m)	Planimetric Error (m)
1	10 m	-1.005	0.141	0.301	1.015
2	20 m	-0.708	0.151	0.285	0.724
3	30 m	-1.175	0.209	0.328	1.193
4	40 m	-1.001	0.194	0.312	1.020
		Mean error (m)		0.307	0.988

Table 3. Difference between calculated and real coordinates – Experiment II.

Comparing these results (Table 3) with those obtained for Experiment I (Table 2), it is observed that, although the altimetric error is slightly higher for Experiment II, the planimetric error is approximately four times lower. This can be explained by the phototriangulation processing results: in Experiment II, phototriangulation was performed using a larger number of images and ground control points, with images captured from different directions – unlike Experiment I, where phototriangulation was based on images acquired from the same direction. As phototriangulation provides the sensor orientation parameters (κ, ϕ, ω) used in monoplotting, its quality directly affects the accuracy of the results.

For the four images from Experiment II under analysis (Figure 7), the horizontal distance between the position of the person defined by the monoplotting procedure and the UAV position is approximately 30 meters. As the variation between these distances is minimal (Table 4), simple linear regression was not applied to the results of Experiment II.

Image	1	2	3	4	Mean
Distance (m)	30.768	30.574	30.194	30.049	30.396

Table 4. Horizontal distance between Person III and the UAV projection on the ground.

Considering the additional image acquired in Experiment II – with a horizontal distance of 70 meters between the UAV and the detected person – the resulting planimetric error was 1.796 meters. This is approximately 1.8 times greater than the average planimetric error for images in which the UAV and Person III were about 30 meters apart. These results do not fit the line equation derived from the values of Experiment I (Figure 9). Therefore, a larger and more representative statistical sample is required to derive an equation capable of mitigating errors associated with the estimated position of the person as a function of the person–UAV distance. Only then can the calculated coordinates be reliably used for real-time UAV path replanning applications.

Lastly, it is important to emphasize that the average processing time per image/person for executing the Monoplotting process is approximately 1.24 seconds when using a NVIDIA GeForce RTX 3060 6 GB GPU. Considering that, for the proposed real-time path replanning method, human detection will be checked along the route every 30 m at a minimum, and that the flight speed of the Mavic 3M will be at most 10 m/s, there will be a 3-second window available for human detection, coordinate estimation, and UAV path replanning. Therefore, it can be concluded that the Monoplotting processing time is sufficient for real-time execution of the method.

4. Challenges and future research

Human detection in UAV imagery using the YOLOv7 model trained on the Unicamp-UAV dataset presented a high false negative rate, i.e., many missed detections – reflected by a low recall value (R = 54.4%). The more recent versions of the YOLO model – YOLOv8, YOLOv9, YOLOv10, YOLO11, and YOLO12 – have shown improved object detection accuracy compared to YOLOv7. Therefore, future research should consider conducting a comparative analysis of the latest YOLO versions to evaluate which model is most suitable for specific applications, such as UAV-based last-mile delivery operations, where a fast and accurate model is essential for ensuring real-time routing safety. A comparative analysis of these models was not conducted in the present study due to the high computational cost of training with the Unicamp-UAV dataset. Training YOLOv7 on the Unicamp-UAV dataset required approximately 67 minutes per epoch using an NVIDIA GeForce RTX 3060 6 GB GPU, with 300 epochs needed to achieve model convergence.

Expanding the Unicamp-UAV dataset by incorporating RGB images with characteristics different from those already included could enhance the generalization capability of the deep learning model and, consequently, improve human detection in UAV imagery. However, it is important to note that such an expansion would further increase the training time of the YOLO models.

Since the Unicamp-UAV dataset comprises only RGB images, data captured by sensors other than the RGB camera, despite the DJI Mavic 3M UAV being equipped with a multispectral camera, were not included. Previous studies, such as Guettala et al. (2022), have demonstrated the effectiveness of using thermal imagery to train deep learning models for human detection. Therefore, future work should investigate the improvements in performance from integrating thermal data for human detection in UAV imagery.

The IMU system has the greatest impact on the accuracy of 3D coordinates obtained via the monoplotting process. Thus, accurate image orientation is essential. Given the high sensitivity of the EOPs related to sensor attitude (κ, ϕ, ω), ideally, a high-precision IMU should be used. However, the associated cost remains a significant challenge. An alternative approach would be to refine the EOPs using image-based techniques such as Simultaneous Localization and Mapping (SLAM). In UAV applications, visual-inertial SLAM, for instance, has shown

excellent performance in terms of robustness and accuracy by integrating IMU measurements (Zhuang et al., 2024).

Nonetheless, this type of approach requires onboard processing due to its real-time characteristics. Thus, the weight and power consumption of the sensors and onboard processing unit, the synchronization of data between sensors and the flight controller, and real-time processing remain challenges. Each of these factors must be carefully addressed for effective UAV operation. Future research should consider refining the UAV's position and orientation based on SLAM, especially in scenarios where GNSS signal loss may occur during flight – conditions that could further increase planimetric errors resulting from the monoplotting process, and which were not accounted for in this study.

5. Final considerations

The YOLOv7 model trained with the Unicamp-UAV dataset demonstrated satisfactory performance for human detection in UAV imagery. Although the model inference occasionally failed to detect individuals in the images (FNs), the person closest to the UAV was consistently and rapidly detected in the analyzed images, ensuring real-time UAV path replanning for last-mile delivery applications. However, challenges associated with human detection in UAV imagery – particularly variations in individual size and pose, as well as image quality and contrast – affected detection confidence in certain scenarios. Furthermore, the proper definition of the confidence threshold during model inference is essential to avoid false positives, especially when inference images are acquired using a UAV different from the one employed to create the training dataset.

Accurate estimation of the detected person's coordinates along the UAV route using the Monoplotting procedure requires careful attention to the sensor's rotation angles. This is particularly important for aircraft equipped with high-precision positioning systems. To mitigate discrepancies between estimated and actual coordinates, mathematical correction strategies may be applied, which require further investigation. Another promising approach involves research on the integration of SLAM techniques for refining EOPs, especially the UAV's rotation angles.

Acknowledgements

The authors thank FAEPEX (Teaching, Research and Extension Support Fund) for supporting this research [grant 2498/24], as well as the Federal Institute of Education, Science and Technology of the South of Minas Gerais (IFSULDEMINAS) for its support and encouragement.

References

Agarwal, A., Ratha, N., Vatsa, M., Singh, R., 2021. Impact of Super-Resolution and Human Identification in Drone Surveillance, in: *2021 IEEE International Workshop on Information Forensics and Security (WIFS)*. Presented at the 2021 IEEE International Workshop on Information Forensics and Security (WIFS), IEEE, Montpellier, France, pp. 1–6. https://doi.org/10.1109/WIFS53200.2021.9648399.

Akshatha, K.R., Karunakar, A.K., Satish Shenoy, B., Phani Pavan, K., Dhareshwar, C.V., Johnson, D.G., 2023. Manipal-UAV person detection dataset: A step towards benchmarking datasets and algorithms for small object detection. *ISPRS Journal of Photogrammetry and Remote Sensing* 195, 77–89. https://doi.org/10.1016/j.isprsjprs.2022.11.008.

Australian Government, 2023. Civil Aviation Safety Regulations 1998.

Bachir, N., Memon, Q., 2022. Investigating YOLOv5 for Search and Rescue Operations Involving UAVs: Investigating YOLO5, in: *2022 The 5th International Conference on Control and Computer Vision*. Presented at the ICCCV 2022: 2022 The 5th International Conference on Control and Computer Vision, ACM, Xiamen, China, pp. 200–204. https://doi.org/10.1145/3561613.3561644.

DECEA, 2023. ICA 100-40 - Aeronaves não Tripuladas e o Acesso ao Espaço Aéreo Brasileiro.

DJI, 2025. DJI Mavic 3 Enterprise. DJI Enterprise. URL https://enterprise.dji.com/mavic-3-enterprise/photo (accessed 6.21.25)

Dousai, N.M.K., Lonearic, S., 2022. Detecting Humans in Search and Rescue Operations Based on Ensemble Learning. *IEEE Access* 10, 26481–26492. https://doi.org/10.1109/ACCESS.2022.3156903.

Fluehler, M., Niederoest, J., Akca, D., 2005. Development of an educational software system for the digital monoplotting, in: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences.* Presented at the ISPRS Workshop Commissions VI/1 – VI/2 : Tools and Techniques for E-Learning, ETH, Eidgenössische Technische Hochschule Zürich, Institute of Geodesy and Photogrammetry. https://doi.org/10.3929/ethz-a-005126679.

Golcarenarenji, G., Martinez-Alpiste, I., Wang, Q., Alcaraz-Calero, J.M., 2021. Efficient Real-Time Human Detection Using Unmanned Aerial Vehicles Optical Imagery. *International Journal of Remote Sensing* 42, 2440–2462. https://doi.org/10.1080/01431161.2020.1862435.

Guettala, W., Sayah, A., Kahloul, L., Tibermacine, A., 2022. Real Time Human Detection by Unmanned Aerial Vehicles, in: *2022 International Symposium on Innovative Informatics of Biskra (ISNIB)*. Presented at the 2022 International Symposium on Innovative Informatics of Biskra (ISNIB), IEEE, Biskra, Algeria, pp. 1–6. https://doi.org/10.1109/ISNIB57382.2022.10075707.

Gupta, H., Verma, O.P., 2022. Monitoring and surveillance of urban road traffic using low-altitude drone images: a deep learning approach. *Multimed Tools Appl* 81, 19683–19703. https://doi.org/10.1007/s11042-021-11146-x.

Kamath, V., Renuka, A., 2023. Deep learning based object detection for resource-constrained devices: Systematic review, future trends, and challenges ahead. *Neurocomputing* 531, 34–60. https://doi.org/10.1016/j.neucom.2023.02.006.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L., 2014. Microsoft COCO: Common Objects in Context, in: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (Eds.), Computer Vision – ECCV 2014, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 740–755. https://doi.org/10.1007/978-3-319-10602-1_48.

Liu, C., Szirányi, T., 2021. Real-Time Human Detection and Gesture Recognition for On-Board UAV Rescue. *Sensors* 21, 2180. https://doi.org/10.3390/s21062180.

Liu, X., Wang, C., Liu, L., 2022. Research on Pedestrian Detection Model and Compression Technology for UAV Images. *Sensors* 22, 9171. https://doi.org/10.3390/s22239171.

Mikhail, E.M., Bethel, J.S., McGlone, J.C., 2001. Introduction to Modern Photogrammetry. John Wiley & Sons.

Othman, N.A., Aydin, I., 2023. Development of a Novel Lightweight CNN Model for Classification of Human Actions in UAV-Captured Videos. *Drones* 7, 148. https://doi.org/10.3390/drones7030148.

Raja, G., Saravanan, G., Dev, K., 2023. 6G-Assisted UAV-Truck Networks: Toward Efficient Essential Services Delivery. *IEEE Communications Standards Magazine* 7, 4–9. https://doi.org/10.1109/MCOMSTD.0003.2200003.

Rapidlasso, 2024. LAStools. rapidlasso GmbH. URL https://rapidlasso.de/product-overview/ (accessed 10.3.24).

Sell, L., 2024. HumanSignal/labelImg.

Serghei, T.-L., Ichim, L., Popescu, D., 2022. Human Detection in Restricted Areas Using Deep Convolutional Neural Networks, in: *2022 30th Telecommunications Forum (TELFOR)*. Presented at the 2022 30th Telecommunications Forum (TELFOR), IEEE, Belgrade, Serbia, pp. 1–4. https://doi.org/10.1109/TELFOR56187.2022.9983720.

Serghei, T.-L., Pârvu, P.V., Serghei, M.-O., Popescu, D., Ichim, L., 2023. Deep Convolutional Neural Networks for Real-Time Human Detection and Tracking on UAVs Embedded Systems, in: *2023 31st Mediterranean Conference on Control and Automation (MED)*. Presented at the 2023 31st Mediterranean Conference on Control and Automation (MED), pp. 311–316. https://doi.org/10.1109/MED59994.2023.10185820.

Simões, D.P., de Oliveira, H.C., dos Santos, R.L., 2023. Analysis of Exterior Orientation Parameters on Monoplotting procedure for avoiding obstacles in UAV real-time routing. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* X-1/W1-2023, 423–429. https://doi.org/10.5194/isprs-annals-X-1-W1-2023-423-2023.

Simões, D.P., Oliveira, H.C., Garcia, M.V.Y., 2022. UAV 3-D Path Planning Based on High-Resolution DSM, DTM, and True Orthomosaic. *IEEE Geoscience and Remote Sensing Letters* 19, 1–5. https://doi.org/10.1109/LGRS.2022.3219733.

Sinha, K.P., Kumar, P., 2023. Human activity recognition from UAV videos using a novel DMLC-CNN model. *Image and Vision Computing* 134, 104674. https://doi.org/10.1016/j.imavis.2023.104674.

Song, H., Song, W., Cheng, L., Wei, Y., Cui, J., 2024. PDD: Post-Disaster Dataset for Human Detection and Performance Evaluation. *IEEE Transactions on Instrumentation and Measurement* 73, 1–14. https://doi.org/10.1109/TIM.2023.3346508.

Sun, T., Chen, H., Duan, X., Lou, H., Liu, H., 2022. Small object detection method based on YOLOv5 improved model, in: 2022 *IEEE 5th International Conference on Information Systems and Computer Aided Education (ICISCAE)*. Presented at the 2022 IEEE 5th International Conference on Information Systems and Computer Aided Education (ICISCAE), pp. 934–940. https://doi.org/10.1109/ICISCAE55891.2022.9927606.

Symeonidis, C., Mademlis, I., Pitas, I., Nikolaidis, N., 2022. Auth-Persons: A Dataset for Detecting Humans in Crowds from Aerial Views, in: *2022 IEEE International Conference on Image Processing (ICIP).* Presented at the 2022 IEEE International Conference on Image Processing (ICIP), IEEE, Bordeaux, France, pp. 596–600. https://doi.org/10.1109/ICIP46576.2022.9897612.

Ultralytics, 2025. Model Prediction with Ultralytics YOLO. Ultralytics YOLO Docs. URL https://docs.ultralytics.com/modes/predict (accessed 11.28.24).

Vijayakumar, A., Vairavasundaram, S., 2024. YOLO-based Object Detection Models: A Review and Their Applications. *Multimed Tools Appl* 1–40. https://doi.org/10.1007/s11042-024-18872-y.

Wang, C.-Y., Bochkovskiy, A., Liao, H.-Y.M., 2022. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. https://doi.org/10.48550/arXiv.2207.02696

Wong, K.-Y., 2022. WongKinYiu/yolov7.

Zhang, A., Lipton, Z.C., Li, M., Smola, A.J., 2023. Dive into Deep Learning. D2L.ai.

Zhang, J., Liang, X., Wang, M., Yang, L., Zhuo, L., 2020. Coarse-to-fine object detection in unmanned aerial vehicle imagery using lightweight convolutional neural network and deep motion saliency. *Neurocomputing* 398, 555–565. https://doi.org/10.1016/j.neucom.2019.03.102.

Zhu, P., Wen, L., Du, D., Bian, X., Fan, H., Hu, Q., Ling, H., 2022. Detection and Tracking Meet Drones Challenge. *IEEE Trans. Pattern Anal. Mach. Intell.* 44, 7380–7399. https://doi.org/10.1109/TPAMI.2021.3119563.

Zhuang, L., Zhong, X., Xu, L., Tian, C., Yu, W., 2024. Visual SLAM for Unmanned Aerial Vehicles: Localization and Perception. *Sensors* 24, 2980. https://doi.org/10.3390/s24102980.