Towards Active Vision with UAVs in Marine Search and Rescue: Analyzing Human Detection at Variable Altitudes

Li Qingqing[1], Jussi Taipalmaa[2], Jorge Peña Queralta[1], Tuan Nguyen Gia[1], Moncef Gabbouj[2], Hannu Tenhunen[1], Jenni Raitoharju[3], Tomi Westerlund[1]

[1]Turku Intelligent Embedded and Robotic Systems Lab, University of Turku, Finland
Emails: [1]{qingqli, jopequ, tunggi, hatenhu, tovewe}@utu.fi

[2]Department of Computing Sciences, Tampere University, Finland
Emails: [2]{jussi.taipalmaa, moncef.gabbouj}@tuni.fi

[3]Programme for Environmental Information, Finnish Environment Institute, Jyväskylä, Finland
Email: jenni.raitoharju@environment.fi

***Abstract*—Unmanned Aerial Vehicles (UAVs) have been playing an increasingly active role in supporting search and rescue (SAR) operations in recent years. The benefits are multiple such as enhanced situational awareness, status assessment, or mapping of the operational area through aerial imagery. Most of these application scenarios require the UAVs to cover a certain area. If the objective is to detect people or other objects, or analyze in detail the area, then there is a trade-off between speed (higher altitude coverage) and perception accuracy (lower altitude). An optimal point in between requires active perception on-board the UAV to dynamically adjust the flight altitude and path planning. As an initial step towards active vision in UAV search in maritime SAR scenarios, in this paper we focus on analyzing how the flight altitude affects the accuracy of object detection algorithms. In particular, we quantify what are the probabilities for false negatives and false positives in human detection at different altitudes. Our results define the correlation between the altitude and the ability of UAVs to effectively detect people in the water.**

***Index Terms*—Active Vision; Flight Altitude; Dynamic Altitude; Object Detection; Human Detection; Marine Search and Rescue (SAR); Unmanned Aerial Vehicles (UAV);**

Fig. 1: Illustration of active-vision-based search in maritime environments with UAVs. A single UAV can first fly higher to cover larger areas and descend in the event of a positive detection to increase reliability. Search time can then be optimized by dynamically adjusting the altitude depending on the perception confidence.

I. Introduction

Recent years have seen an increasingly wider adoption of unmanned aerial vehicles (UAVs) to support search and rescue (SAR) operations. Owing to their fast deployment, speed and aerial point of view, UAVs can aid quick response teams, but also in longer-term monitoring and surveillance [1]. Some of the main applications of UAVs in these scenarios are real-time mapping of the operational area or delivery of emergency supplies. In particular, UAVs can bring a significant increase of the response team's situational awareness and detect objects and people from the air, specially those in need of rescue [2]. An overview of recent research in this area is available in [3], where UAVs for SAR operations are characterized based on the operational environment, the type of robotic systems in use, and the onboard sensing capabilities of the UAVs.

We are interested in optimizing the support that UAVs can provide in maritime SAR operations (see Fig. 1), but also for monitoring and surveillance in maritime environments, where they have already been widely utilized [4]. Maritime SAR operations might occur in both normal and harsh environments. For example, according to the Spanish national drowning report [5], in 2019 over 40% of drownings happened on a beach, around 60% of the incidents happened between 10:00 and 18:00, and in 20% of the cases lifeguards were present in the area. Therefore, there is still a need for better solutions for monitoring and supporting SAR operations in safeguarded beaches, lakes or rivers even with favorable weather conditions, which can then be extended towards rougher environments as the technology evolves. In this paper, we study the detection of people in mostly still waters at different altitudes. In the future, we aim to utilize this information within an active vision algorithm that can dynamically adapt the flight plan of UAVs towards optimization of search speed and reliability.

In terms of UAV-based perception, deep learning (DL) methods have become the de-facto standards in object detection and image segmentation with great success across multiple domains [6], [7]. In this paper, we utilize the YOLOv3 [6] architecture and characterize its performance for human detection on still water surfaces. Within the machine perception field, active vision has been a topic of interest that has gained increasing research interest, owing to the multiple advances in DL and accessibility of UAV platforms for research. Active vision has been successfully applied for single and multi-agent tracking [8], but we have observed a gap in the literature in terms of active vision for search and area coverage. The most active research direction in active perception is currently reinforcement learning (RL) [9]. However, we consider in this paper a more traditional approach. An RL approach can be challenging owing to the lack of realistic simulators to train models for sea SAR.

Deep learning for perception in maritime environments is limited by the lack of realistic training datasets openly available. Moreover, a key challenge for UAV-based person search and detection in these environments is the relatively small size of objects to be detected in comparatively large areas to be searched [10]. There is an evident trade-off between speed and area coverage, and reliability of both positive and negative detection. An additional challenge is that the view of people at sea from the air is only partial, as a significant portion of the body is immersed in the water. Water reflection and refraction effects might also distort the shape. In order to train YOLOv3 to adapt to this scenario, and owing to the lack of open data for detecting people in water, we collected over 450 high-resolution images to train, validate and test our model. The images have been taken at altitudes ranging from 20 m to 120 m.

This is, to the best of our knowledge, the first paper to analyze the perception accuracy for UAVs with RGB cameras in maritime environments as a function of their altitude. The results can be generalized by accounting for the size in pixels of the persons to be detected assuming well-focused images. Moreover, the retrained YOLO model outperforms the state-of-the-art in object classification, as it has been trained to detect people even when only their head emerges above the water level. The retrained YOLO model can be applied for people swimming but also standing near the shore in a beach.

The rest of the paper is organized as follows. Section II briefly overviews previous research in active vision, on one side, and maritime SAR operations supported by UAVs, on the other. We then describe the main objectives of our study in Section III, together with data acquisition and model training details. Section IV reports our experimental results and Section V concludes the work.

II. Background

Multiple works have demonstrated the benefits of integrating UAVs to maritime SAR operations [11], [12]. Typical sensors onboard UAVs are RGB, RGB-D and thermal cameras, 3D lidars, and inertial/positional sensors for GNSS and altitude estimation [13], [14]. With these sensors, UAVs can aid in SAR operations by mapping the environment, locating victims and survivors, and recognising and classifying different objects [13]. From the perception point of view, DL methods have become the predominant solution for detecting humans or other objects [7], [15], [16].

Human detection is a sub-task of object detection that is of particular interest for SAR robotics [17]. Some of the most popular neural network architectures for object detection are R-CNN [18], Fast-RCNN [19], and YOLO [6]. In particular, YOLOv3 is the current state-of-the-art for real-time detection, able of fast inference and high accuracy [6]. In this paper, we re-train the YOLOv3 network with a new dataset for detecting people in the water.

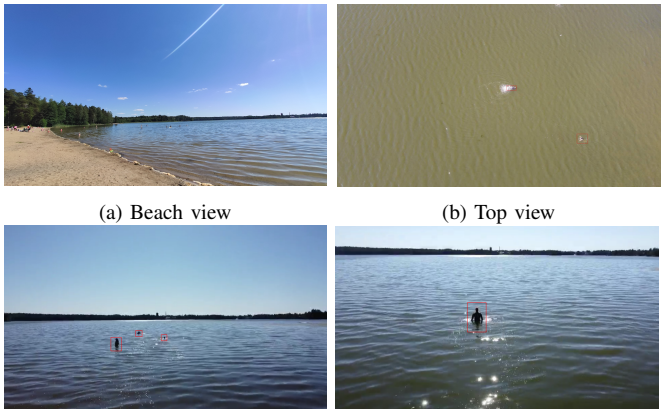
Active perception has been defined as:

> An agent is an active perceiver if it knows *why* it wishes to sense, and then chooses *what* to perceive, and determines *how*, *when*, and *where* to achieve that perception. [20]

In UAV-aided maritime SAR operations, algorithms for area coverage and human search incorporating active vision need to be aware that their main objective is to find humans (why), and need to be able to dynamically adjust their path planning and orientation to achieve higher-confidence results (what). This latter aspect can be achieved by, for instance, adjusting their height and camera pitch, or by moving around the person to get a better angle (how, where and when).

Active vision has been increasingly adopted in different object detection tasks. However, no previous research has, to the best of our knowledge, focused on active vision for detection of humans in SAR scenarios. We therefore list here some other relevant works in the area. Ammirato et al. presented a dataset for robotic vision tasks in indoor environments using RGBD cameras with the introduction of an active vision strategy using Deep RL to predict the next best move for object detection [21]. Juan et al. presented an autonomous Sequential Decision Process (SDP) for active perception of targets in uncertain and cluttered environments, with experiments conducted in a simulated SAR scenario [22]. Davide et al. applied active vision to a path planing algorithms that enabled quadrotor flight through narrow gaps in indoor complex environments [23]. Manuela et al. applied bio-inspired active vision for object avoidance with wheel robots in indoor environments [24]. In SAR operations, once a target has been identified, continuously updating the position of target is essential, so that path planning for the rescue teams can be adjusted. This can be achieved though active tracking [25].

In terms of detecting people in maritime environments, Eleftherios et al. presented a real-time human detection system using DL models that run on-board UAVs to detect open water swimmers [26]. The authors, however, do not study the accuracy of the perception for different altitudes or positions. In this work, we focus on analyzing human detection as a trade-off between larger area coverage (higher altitude) and higher amount of detail in the images (lower altitudes).



(a) Beach view (b) Top view

(c) Low altitude (d) Back light

Fig. 2: (a) Example images of terrain at Littoinen Lake, Finland, (b) The top view of swimmer, (c) The far view of swimmers, (d) The close view of swimmer

In general, we see a clear trend towards a more widespread utilization of UAVs in SAR missions and DL models for perception (either onboard or offloading computation). We have found, however, no previous works exploring the correlation between the altitude at which UAVs fly and the detection accuracy in maritime SAR scenarios.

III. Methodology

This section describes the data and details of the training process for the perception algorithm. We also outline the metrics that are analyzed in our experiments.

A. Data Acquisition

Owing to the lack of labeled open data showing people in water, and in particular data labeled with the flight altitude, we have collected data from people swimming and standing in a lake. The dataset contains 458 labeled photos that are taken by the camera mounted on the UAV. The camera has a fixed focal length of 24 mm (35 mm format equivalent) with a field of view of 83°and an aperture f/2.8. The images have a resolution of 9 MP (4000 by 2250 pixels), and were recorded near the beach area of Littoistenjärvi Lake (60.4582 N, 22.3766 E), shown in Fig. 2 (a), Turku, Finland.

Each photo captures one or more people that are either swimming or standing in the lake at different heights and angles. Some examples are shown in Fig. 2 (b), (c) and (d). However, the majority of pictures were taken with a gimbal pitch of -90°(top-view images). The dataset contains 2D bounding boxes for two classes: *persons* and *other objects*, the latter one being used for animals in the water and other floating objects. In addition to the bounding boxes, each image contains information about the GPS position, relative altitude to the take-off point (just above the water level), and pitch angle of the camera gimbal (from horizontal images with 0°pitch to top-view images with -90°pitch). The relative altitude ranges from 0 m to 143 m. While the dataset has been acquired with good weather conditions and mostly still waters, variable light conditions are also introduced. This results in different colors for both water and people, as can be seen in Fig. 2 (b) and (d). Some of the swimmers use swimming caps of different colors and wear different types of swimming suits.

B. Training and test setup

Training and testing were done with the YOLOv3 real-time object detection model [6]. The YOLOv3 model pre-trained with ImageNet [27] was trained again with our dataset using transfer learning. Training is done in a way where all but the last three layers are frozen for the first 50 epochs and then unfrozen and trained further for another 50 epochs with batch size of 32 and learning rate of 0.001.

Each image contains between 1 and 50 object instances. The objects are divided into two classes: *'person'*, containing 2454 instances, and *'something else'*, containing 238 instances, mostly birds but also some other objects floating in the water. All the images were labeled manually, using bounding boxes with the Labelbox annotation tool [28]. Training and testing were done using 4-fold cross-validation, randomly splitting the images using a 75/25 train/test split. We refer to the re-trained model as the task-specific model hereinafter.

C. Metrics

Object detection performance was evaluated using PASCAL VOC challenge metrics [29] provided by [30]. We calculated average perception (*AP*) for both classes separately and mean average perception (*mAP*) over both classes using different intersection over union (*IoU*) thresholds. The comparison in performance was done between the pre-trained YOLOv3 model and the task-specific model with our data using transfer learning. Furthermore, since our objective is to analyze the correlation between the performance of the human detection and the altitude, we also analyze how the detection confidence and the ratio of false positives and false negatives changes as a function of the altitude.

IV. Experimental Results

In this section, we assess the performance of the trained model as a classifier using the mean average precision for different IoU thresholds, but also its usability for active-vision-based control where the input to the algorithm is the confidence of the model on each of its detections.

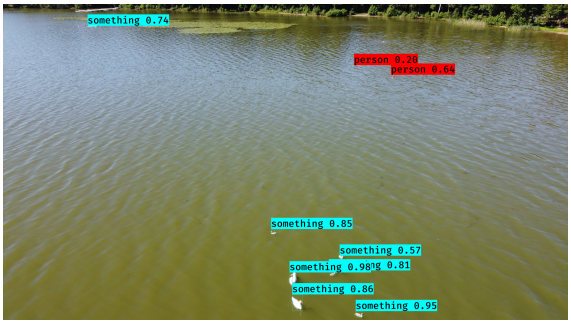
Some representative example detections made by the task-specific model are illustrated in Fig. 3. In Fig. 3a, we observe how the network is able to pinpoint the location of people in the image, but the bounding box appears around the turbulent water rather than around the person itself. However, not all objects or turbulent areas are detected as people, as other objects are also properly identified (Fig. 3b. In Fig. 3b, we also observe that people can be located far away when the gimbal pitch is closer to 0°. Finally, we see that even at high altitudes, the confidence remains high and people are detected also when immersed (Fig. 3c).

The performance of the task-specific model compared to the pre-trained YOLOv3 network is shown in Table I, where we see that the task-specific model is clearly superior. In

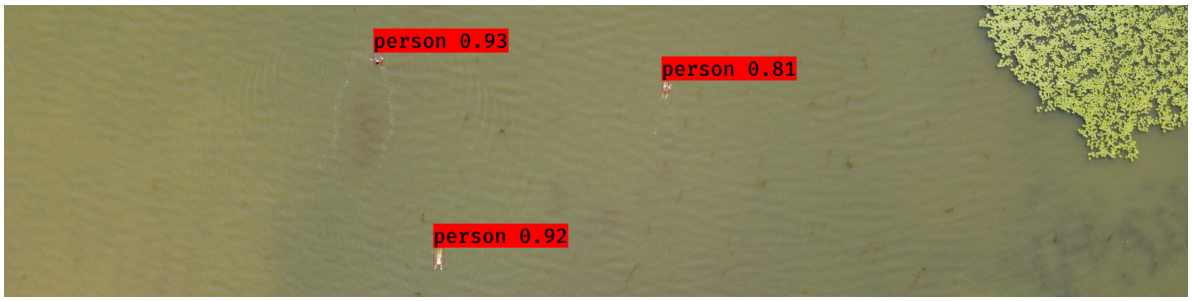
TABLE I: mAP-scores for different IoU-thresholds.

		IoU-threshold									
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Model	**Task-specific model**	0.6985	0.6984	0.6972	0.6954	0.6934	0.6883	0.6780	0.6384	0.5422	0.0000
	Pre-trained YOLOv3	0.0547	0.0533	0.0533	0.0528	0.0514	0.0514	0.0514	0.0507	0.0440	0.0000

(a) Detection of one person (high confidence), and turbulent water next to another (lower confidence). Altitude: 37 m. Pitch: -80°.



(b) Detection of other objects but missing two persons in the distance. Altitude: 12 m. Pitch: -25°.



(c) Successful detection of three people at high altitude, one of them fully immersed in the water (only a portion of the original image is shown). Altitude: 86 m. Gimbal pitch: -90°.

Fig. 3: Samples of detections made using the task-specific model.

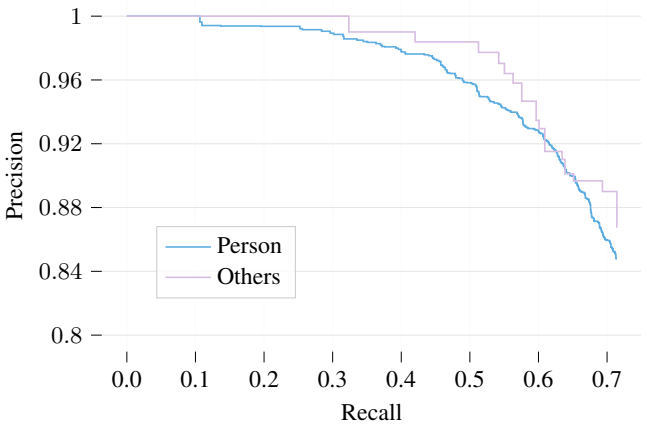


Fig. 4: Precision x Recall curve for class 'person' and 'something else' using IoU-threshold 0.5 with the task-specific model.

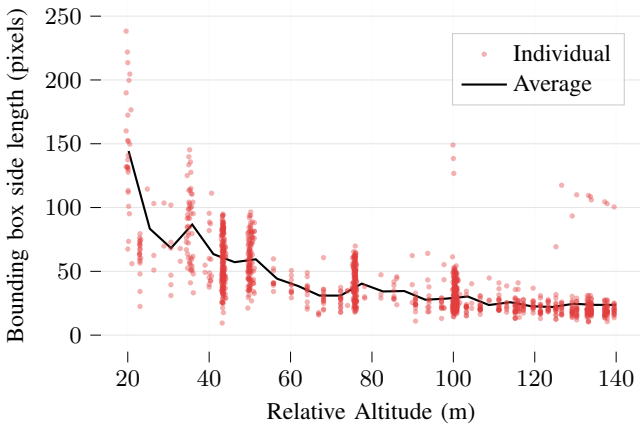
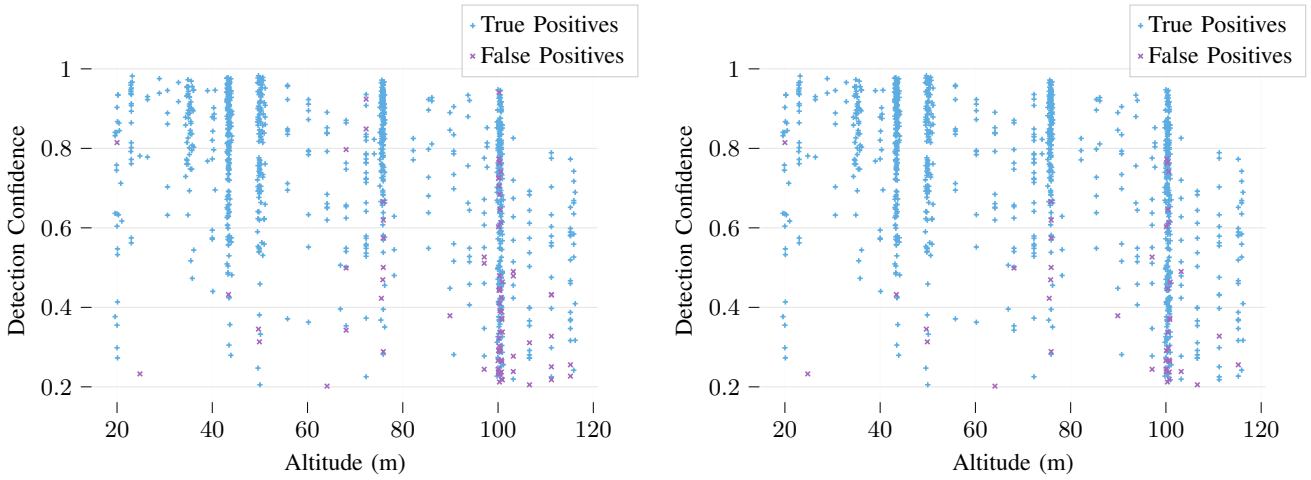


Fig. 5: Side length of the ground truth bouding boxes, in pixels, based on the altitude.

terms of the precision × recall curves, those corresponding to classes *'person'* and *'something else'* are provided in Fig. 4.

Next, we analyze performance at different altitudes. The significance of the altitude is, however, relative to the resolution of the camera and its ability to produce clear images. The camera pitch is also important as illustrated. In order to provide results that are more generalizable, Fig. 5 shows the size in pixels of the ground truth bounding boxes.

Fig. 6a shows all the person detections plotted in terms of their confidence against the altitude, using $IoU = 0.1$ to



(a) Confidence with IOU

(b) Confidence with DIST

Fig. 6: Confidence of individual detections as a function of the relative UAV altitude. We observe a clear difference between high-confidence and true positives under the threshold of 100 m, with lower confidence and higher rate of false positives above it.

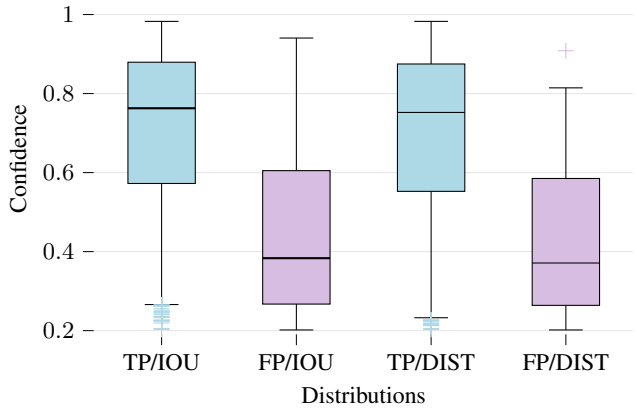


Fig. 7: Distributions for the confidence of true positive (TP) and false positive (FP) detections (DIST and $IoU = 0.1$).

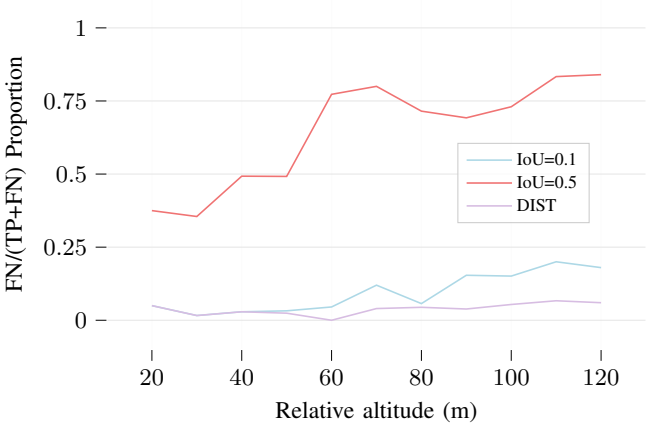


Fig. 8: Proportion of false negatives (FN) over true positives (TP) and FN. This gives an idea of the probability of missing a person.

consider true positives. We have set the IoU to 0.1 because we are only interested in pointing to the approximate location of persons but not their exact size and place. For altitudes under 60 m, over 98.8% of the detections with a confidence above 0.5 are correct. A clear threshold appears at an altitude of 90 m. Above 90 m, 83.3% of the detections are correct.

In some of the test images, we have noticed that the model detects turbulence in the water created by people as persons, and not the full bodies of the people themselves. Because we are not interested in analyzing how capable the task-specific model is of generating accurate bounding boxes, but instead on pointing to the approximate location of people at sea, we might also want to consider as correct detection boxes that are just adjacent to actual people. In Fig. 6b, we have plotted the confidence as a function of the altitude, but now using a distance in pixels of less than 100 between the ground truth and the predicted box (DIST) to assume that a detection is correct. We now see that all except one of the positive detections with a confidence of over 70% are correct for an altitude up to 100 m. For a confidence above 45%, all but one detections are correct up to an altitude of 70 m. The distributions of the true positives and false positives for each of the two metrics (IoU, DIST) are shown in Fig. 7. There is a clear threshold just under a confidence of 0.6, with almost 75% of true positive having a confidence over 0.6, and almost 75% of false positives having a lower confidence.

In order to evaluate this model within its context for SAR missions, we also need to take into account that false positives do not necessarily have a significant impact on the search performance, but false negatives do, as they mean that the UAV misses a person. We have therefore plotted in Fig. 8 the proportion of false negatives over true positives. If we use the pixel distance to consider a detection as correct, then the proportion remains under 10% for all altitudes. With $IoU = 0.5$, however, over 50% of the people in the water are undetected. However, we do not consider this an effective way of evaluating a detection in this scenario.

V. Conclusion

With UAVs increasingly penetrating multiple civil domains and, among them, search and rescue operations, more complex control mechanisms are required for more autonomous UAVs. To that end, active perception is one of the most promising research directions. In UAV search, active vision can be exploited to optimize the flight plan based on the confidence of the DL vision algorithms. We have presented preliminary work that studies the confidence of a re-trained YOLOv3 model for detecting people in the water for altitudes ranging from 20 m to 120 m. With a custom dataset, we have seen a major performance increase with respect to the pre-trained YOLOv3 network. Our results show a clear correlation between the altitude and the confidence of the detections and between the confidence and the correctness of the detections. When considering as true positives detections near actual people (e.g., over water turbulence created by people), we have seen that the proportion of false negatives remains low even for high altitudes, and the proportion of false positives over true positives drops significantly for all predictions with a confidence over 60%. Finally, we have observed a clear altitude threshold at around 100 m after which confidence and accuracy drop.

The results presented in this paper will serve as the starting point towards the design of active-vision-based search with UAVs in marine SAR operations. In future works, we will also incorporate the camera pitch into the analysis. The dataset will be made publicly available with further additions.

Acknowledgements

This work was supported by the Academy of Finland's AutoSOS project with grant number 328755.

References

[1] H. Shakhatreh, A. H. Sawalmeh, A. Al-Fuqaha, Z. Dou, E. Almaita, I. Khalil, N. S. Othman, A. Khreishah, and M. Guizani, "Unmanned aerial vehicles (uavs): A survey on civil applications and key research challenges," *IEEE Access*, vol. 7, pp. 48 572–48 634, 2019.

[2] J. Peña Queralta, J. Taipalmaa, B. C. Pullinen, V. K. Sarker, T. N. Gia, H. Tenhunen, M. Gabbouj, J. Raitoharju, and T. Westerlund, "Collaborative multi-robot search and rescue: Coordination and perception," *arXiv preprint arXiv:2008.12610 [cs.RO]*, 2020.

[3] S. Grogan, R. Pellerin, and M. Gamache, "The use of unmanned aerial vehicles and drones in search and rescue operations–a survey," *Proceedings of the PROLOG*, 2018.

[4] W. Roberts, K. Griendling, A. Gray, and D. Mavris, "Unmanned vehicle collaboration research environment for maritime search and rescue," in *30th Congress of the International Council of the Aeronautical Sciences*. International Council of the Aeronautical Sciences (ICAS) Bonn, Germany, 2016.

[5] Royal Spanish Federation of First Aid and Rescue, "National Drownings Report - Informe Nacional de Ahogamientos (INA)," 2019.

[6] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," *arXiv*, 2018.

[7] S.-J. Hong, Y. Han, S.-Y. Kim, A.-Y. Lee, and G. Kim, "Application of deep-learning methods to bird detection using unmanned aerial vehicle imagery," *Sensors*, vol. 19, no. 7, p. 1651, 2019.

[8] R. Tallamraju, E. Price, R. Ludwig, K. Karlapalem, H. H. Bülthoff, M. J. Black, and A. Ahmad, "Active perception based formation control for multiple aerial vehicles," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 4491–4498, 2019.

[9] D. Gallos and F. Ferrie, "Active vision in the era of convolutional neural networks," in *2019 16th Conference on Computer and Robot Vision (CRV)*, 2019, pp. 81–88.

[10] J. Peña Queralta, J. Raitoharju, T. N. Gia, N. Passalis, and T. Westerlund, "Autosos: Towards multi-uav systems supporting maritime search and rescue with lightweight ai and edge computing," *arXiv preprint arXiv:2005.03409*, 2020.

[11] A. Matos, A. Martins, A. Dias, B. Ferreira, J. M. Almeida, H. Ferreira, G. Amaral, A. Figueiredo, R. Almeida, and F. Silva, "Multiple robot operations for maritime search and rescue in eurathlon 2015 competition," in *OCEANS 2016-Shanghai*. IEEE, 2016, pp. 1–7.

[12] J. Güldenring, L. Koring, P. Gorczak, and C. Wietfeld, "Heterogeneous multilink aggregation for reliable uav communication in maritime search and rescue missions," in *2019 International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*. IEEE, 2019, pp. 215–220.

[13] R. Konrad, D. Serrano, and P. Strupler, "Unmanned aerial systems," *Search and Rescue Robotics—From Theory to Practice*, pp. 37–52, 2017.

[14] H. Surmann, R. Worst, T. Buschmann, A. Leinweber, A. Schmitz, G. Senkowski, and N. Goddemeier, "Integration of uavs in urban search and rescue missions," in *2019 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*. IEEE, 2019, pp. 203–209.

[15] T. Giitsidis, E. G. Karakasis, A. Gasteratos, and G. C. Sirakoulis, "Human and fire detection from high altitude uav images," in *2015 23rd Euromicro International Conference on Parallel, Distributed, and Network-Based Processing*. IEEE, 2015, pp. 309–315.

[16] S. Yong and Y. Yeong, "Human object detection in forest with deep learning based on drone's vision," in *2018 4th International Conference on Computer and Information Sciences (ICCOINS)*. IEEE, 2018, pp. 1–5.

[17] "Autonomous human detection system mounted on a drone," *2019 International Conference on Wireless Communications, Signal Processing and Networking, WiSPNET 2019*, pp. 335–338, 2019.

[18] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580–587.

[19] R. Girshick, "Fast r-cnn," in *International Conference on Computer Vision (ICCV)*, 2015.

[20] R. Bajcsy, Y. Aloimonos, and J. Tsotsos, "Revisiting active perception," *Autonomous Robots*, vol. 42, pp. 177—-196, 2018.

[21] P. Ammirato, P. Poirson, E. Park, J. Košecká, and A. C. Berg, "A dataset for developing and benchmarking active vision," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 1378–1385.

[22] J. Sandino, F. Vanegas, F. González, and F. Maire, "Autonomous uav navigation for active perception of targets in uncertain and cluttered environments," in *2020 IEEE Aerospace Conference*, 2020.

[23] D. Falanga, E. Mueggler, M. Faessler, and D. Scaramuzza, "Aggressive quadrotor flight through narrow gaps with onboard sensing and computing using active vision," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 5774–5781.

[24] M. Chessa, S. Murgia, L. Nardelli, S. P. Sabatini, and F. Solari, "Bio-inspired active vision for obstacle avoidance," in *Conference on Computer Graphics Theory and Applications*, 2014, pp. 1–8.

[25] F. Zhong, P. Sun, W. Luo, T. Yan, and Y. Wang, "AD-VAT: An asymmetric dueling mechanism for learning visual active tracking," in *International Conference on Learning Representations*, 2019.

[26] E. Lygouras, N. Santavas, A. Taitzoglou, K. Tarchanidis, A. Mitropoulos, and A. Gasteratos, "Unsupervised human detection with an embedded vision system on a fully autonomous uav for search and rescue operations," *Sensors*, vol. 19, no. 16, p. 3542, 2019.

[27] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in *CVPR09*, 2009.

[28] Labelbox, "Labelbox," 2019, [Online]. Available: https://labelbox.com.

[29] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, Jun. 2010.

[30] R. Padilla, S. L. Netto, and E. A. B. da Silva, "A survey on performance metrics for object-detection algorithms," in *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, 2020, pp. 237–242.