APH-Yolov7t: Yolov7-Tiny with Attention Prediction Head for Person Detection on Drone-Captured Search and Rescue Scenarios

Vamshi Kodipaka, Lino Forte Marques, Helder Jesus Araújo and Rui Pedro Duarte Cortesão

October 27, 2023

APH-YOLOv7t: YOLOv7-tiny with Attention Prediction Head for Person Detection on Drone-Captured Search and Rescue Scenarios

Vamshi Kodipaka[1,2], Lino Marques[1,2] Helder Araujo[1,2], and Rui Cortesao[1,2]

[1] Institute of Systems and Robotics, ISR-Coimbra
[2] Department of Electrical and Computer Engineering - FCTUC,
University of Coimbra, Portugal
{vamshi.kodipaka, lino, helder, cortesao}@isr.uc.pt

Abstract. Inspection and intervention by drones in rescue operations have growing attention due to multiple causes, including natural and man-related events. Additionally, the rapid advancements in vision sensors, object detection models, and AI-based methods can boost the success of rescue scenarios. Drone navigation involves object scale variations creating a computation load for the scene urge high-speed processing. To solve the two issues mentioned above, we propose the APH-YOLOv7t method that follows Holdout method. In this paper, we introduce a new Attention-based Prediction Head for YOLOv7-tiny. We also present the evaluation results of YOLOv7 the state-of-the-art convolutional neural network-based solution, here is used for robust object detection. In this context of drone navigation there is a need to perform detection of persons on land and sea surfaces allowing to reduce disaster, distress, identify and rescue them. Despite the higher success rate of object detection models, vision complexities make detection tasks on drone-captured images more challenging and this area remains under-explored. We used the existing three search and rescue datasets which are images acquired from drones specific to our objective. Results show that our APH-YOLOv7t method was the most robust attention-based YOLO and comprehensive person detection method for our application, demonstrating a consistently high level of performance in comparison to YOLOv7-tiny. Evaluation results on all three datasets are reported. With this solution, and conditional performance, we demonstrate to be able to satisfy our requirements of a mean average precision (mAP50) of over 0.80 for the person class and operational performance with over 125fps on a single GPU Nvidia RTX2080Ti.

Keywords: person class, person detection, deep learning, fine-tuning, evaluation, Heridal, Mobdrone, SeaDronesSee, attention head, YOLOv7

1 Introduction

Natural and technological disasters, such as hurricanes, earthquakes, eruptions of volcanoes, landslides, and debris flow, wild-land and urban–interface fires, floods, oil spills, and space–weather storms, impose a significant socio-economic burden. It should be imperative to devise vision-based drone solutions that allow rescuers to engage safely while speeding up search operations. One of the most effective measures through drone-based rescue missions is to foster target detection and active monitoring to prevent or respond to imminent danger and risk reduction. Although there are key advances in fully autonomous drone navigation solutions[29][32][33] for Search and Rescue (SAR) missions following crucial principles like selective search pattern, sorted search sweep clutter, and robust data acquisition; we notice that object detection is not mature and still requires further developments. This is because of the crucial vision complexities challenges imposed while capturing images from drones such as pose and scale variations, adverse weather conditions like the presence of snow, dust, fog, low visibility, altitude, and illumination, presence of artifacts like people wearing hats, camouflaged environment with trees and rocks, motion blur and the high image resolution. Thus, developing robust and reliable vision algorithms are of special interest.

Robotic Perception for drones operating in outdoor[34][35] natural environments has been studied for several decades. For drones operating in rescue scenarios, in particular, there is research[36] since the late 80s. Nevertheless, despite many years of research, as described in surveys over time, a substantial amount of problems have yet to be robustly solved. The best examples of the use of drones are after Hurricane Katrina[1], Christchurch earthquake[2], and Paris cathedral firebreak[3] incidents. Thus, drone navigation and its application domain have an unquestionable but impeccable impact on our society. The proposed experiments will contribute to rescue operations by reducing land and sea-surface hazards.

In this paper, we propose an attention-based model, APH-YOLOv7t based on YOLOv7-tiny to solve the problems like object scale variations, and densely packed objects in search and rescue scenarios. The overview of the detection pipeline using APH-YOLOv7t is shown in fig.2,3. We use YOLO-tiny's feature extraction layers with LeakyReLU activations and follow APH-YOLOv7t, which is the original version of the attention head. Totally, APH-YOLOv7t contains four attention-based detection heads separately used for the detection of tiny, small, medium, and large persons. We adopt Convolutional Block Attention Module (CBAM [4]) to sequentially generate the attention map along channel-wise and spatial-wise dimensions as similar to TPH-YOLOv5[5] method. Compared to YOLOv7 and YOLOv7-tiny, APH-YOLOv7t can better deal with drone-captured images. Our contributions are listed as follows:

- We integrate the Attention Prediction Head into YOLOv7-tiny, which can accurately localize objects in high-density scenes.
- We integrate CBAM into YOLOv7-tiny, which helps to find regions of interest in images that have wide coverage.
- Evaluating YOLOv7 and YOLOv7-tiny on existing Heridal[6] Mobdrones[7] and SeaDronesSee[8] datasets which are drone-captured RGB images.

– Scaling and Data-Augmentation such as brightness, saturation, exposure applied on Heridal data to test the detection accuracy.

This paper is organized as follows. Section II of the paper describes the object detection technology and gives a brief collection of SAR datasets. We introduced APH-YOLOv7t and evaluated the performance in the context of a Yolo which we have described in section III. The qualitative and quantitative results are been emphasized in section IV. The conclusion and future work are described as discussion in section V.

2 Related Work

Salient object detection is a computer vision task that involves identifying the most visually distinct or prominent objects in an image or video. The goal of salient object detection is to locate the most important objects in an image, which can then be used for a variety of applications such as image editing, object recognition, and content-aware image resizing. The salient objects in an image are typically defined as the objects that stand out from their surroundings in terms of color, texture, or shape. Here, we used YOLOv7 that typically uses features such as brightness, exposure, saturation and size to identify persons. The different approaches of salient object detection are: feature-based methods[9], region-based[10], deep learning-based methods[11], and hybrid methods[12].

Methods for object detection are a natural progression from non-neural approaches (the approaches to first define features, then use a technique such as a support vector machine[13] to do the classification), to the neural techniques that are able to do end-to-end object detection without specifically defining features and are typically based CNNs which in turn intricate robotic systems. Two-stage CNN detectors[14][15] first propose a set of regions of interest by select search or regional proposal network. The proposed regions are sparse as the potential bounding box candidates can be infinite. Then a classifier only processes region entities. Single-stage detectors[16][17][18] skip the region proposal stage and run detection directly over a dense sampling of possible locations. Two-stage detectors have high localization and recognition accuracy but are slow whereas one-stage detectors have high inference speeds.

This method[19] addressed the problem of comparing the accuracy of human detection in aerial images taken by unmanned aerial systems in SAR missions between an algorithm based on deep neural networks and a SAR expert. An already-existing Heridal[6] image database, with 500 labeled, full-size 4000×3000-pixel real-world images that all contain at least one person are stored. This experiment proved the effectiveness of image processing algorithms as support to SAR missions but failed to evaluate the object detection metrics exclusively. A methodology[20] based on the object detector YOLOv5 is introduced by improving the performances in detecting small objects such as persons in aerial images are evaluated. These algorithms implement shallow layers of the feature extractor to increase the spatial-rich features and help the detector to find small objects. This paper[21] presents the results of an experimental evaluation in which the lightweight version of the YOLOv5 detection, detects humans in danger using two new benchmark datasets specifically designed for SAR with drones. But no pre-processing and data augmentation on input images were investigated.

However, this method[22] proposed a change in bounding box sizes by some percentage of its width and height. To investigate the possible effect of the dataset labeling quality the augmented and non-augmented dataset versions were prepared, leading to significant improvement of performance by loss and mean average precision (mAP) that can be observed in both versions in comparison to experiments without data augmentation. But this method does not investigate the evaluation results on the YOLOv7, a real-time convolutional network for object detection. However, their experiment has limited evaluation results on drone images which load almost double the computation cost as this method uses two models. Attention-based detection heads when integrated into object detection models, are useful for several reasons: improved object localization, scale and context awareness, reduced false positives, inter-object relationships, adaptability, saliency and visualization, robustness to clutter. TPH-YOLOv5[5] also uses an attention mechanism with the CBAM[4] module; two sequential submodules are used to refine feature maps that go through CBAM. However, TPH-YOLOv5 has got the poor inference time that is in few hundreds of milli-seconds, which makes it unsuitable for the real-time applications. We draw inspiration from this method to make YOLOv7 run in less inference time maintaining the accuracy.

There are not abundant benchmark datasets and CNNs models used for SAR operations with drones. We have noticed Heridal[6] (Land), Mobdrone[7] (Sea-surface), SeaDronesSee[8] (Sea-surface), FloodNet[23], Auvsi-Suas[24] (Synthetic characters and shapes), SARD[25] (forestry), and Lacmus[26] (forestry), for carrying out robotic operations for various computer vision tasks. They were already trained and tested using a set of state-of-the-art CNNs but not using YOLOv7. So, as part of considering SAR datasets, our method APH-YOLOv7t along with YOLOv7[27] is trained and evaluated on the Heridal, Mobdrone and SeaDronesSee datasets.

3 Methodology

3.1 Preprocessing the data and YOLOv7

For our application, we are interested in persons by detecting bounding boxes only considering the person class. In real-time the image inputs for our solution consist of three image streams conveyed by an RGB camera, namely Red (R), Green (G), and Blue (B), a frame rate that imposes a requirement on the execution speed for inference to be under 10ms. Another particular requirement for our application is to maximize mAP, a conservative approach that ensures the search activities are effectively conducted. To this end, we evaluated the performance of the state-of-the-art neural network YOLOv7 together on the drone data benchmarks. The image quality and limited image resolution of drone images can be affected by factors such as altitude, lighting conditions, and motion blur. This can make it difficult to identify and locate small objects or subtle features in the image, especially if the images are captured at high altitudes causing obscured scenarios. Drone images usually have a high resolution, thereby pre-processing steps are required to overcome the drawback. In our proposed method, we generated new datasets based on the existing SAR datasets. By pre-processing the drone data[28], we overcome the problem of how to train huge drone images under limited computational resources and find accurate bounding boxes during test time. Data augmentation is applied on Heridal data, that is saturation, brightness, and exposure of -25 to +25 percent are varied.

The generated datasets are then trained on YOLOv7. YOLOv7[27] which is one of the latest advancements among the object detection models which has Extended Efficient Layer Aggregation Layers (EELAN), Model Scaling for concatenation-based models, and Trainable Bag-of-Freebies. It focuses on the number of parameters, and computational density of a model, generate models of different scales and merges multiple computing modules. Thus, the above factors confirm that YOLOv7 is an excellent design for drone datasets. Our proposed experimentation proved the same and attained a real-time performance. The flow chart in fig.1 describes the training and testing pipeline with YOLOv7.

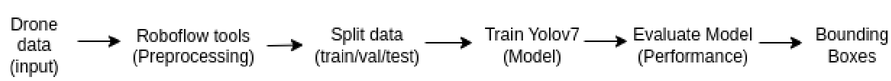

Fig. 1: Simple process flow followed in this work - Train and Test with YOLOv7

3.2 APH-YOLOv7t with CBAM

The core Convolutional Block Attention Module (CBAM[4]) is a simple but effective attention module. It is used in building the Attention-based Prediction Head (APH). Its lightweight modularity allows it to integrate into most of the famous CNN architectures and allows them to train in an end-to-end fashion. In order to accomplish adaptive feature refinement, CBAM multiplies the attention map with the input feature map after progressively inferring it along the two distinct channel and spatial dimensions. The structure of the CBAM module is emphasized in fig. 2. According to the experiment in the study[4], at different scales, the performance of the model significantly increased after integrating CBAM into several models on various classification and detection datasets, demonstrating the efficacy of this module. To enable APH-YOLOv7t to resist the distracting input and concentrate on beneficial target objects, CBAM is used to extract the attention region. To exhibit this, we introduce four CBAM modules which act as detection heads at four varied feature scales. Thus, this leads to a new method of attention heads applied over YOLOv7-tiny benefiting both from reduced model size and an attention mechanism. CBAM is illustrated in fig.2. The detection head of YOLOv7-tiny with CBAM (APH-YOLOv7t) is illustrated in fig.3.

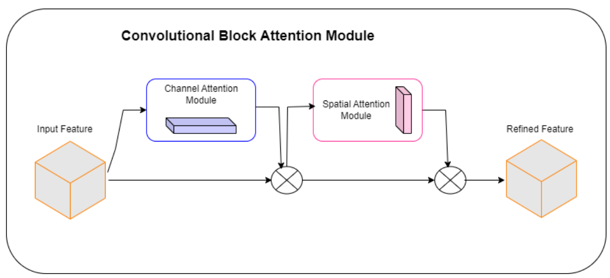

Fig. 2: Convolutional block attention module (CBAM)

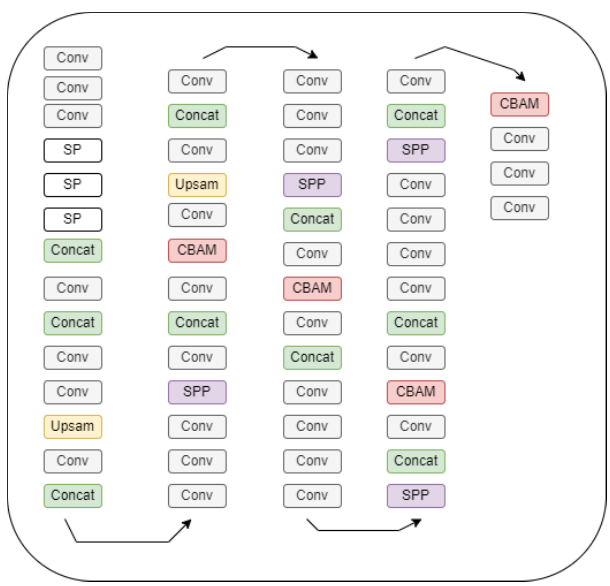

Fig. 3: Detection Head of APH-YOLOv7t with CBAM modules(red). Changes are in red (added CBAM to YOLO-tiny's detection head)

4 Experimental Results

In the description that follows, single-person class ground truth labels are annotated in YOLO format using Roboflow[30] pre-built online annotations tools on all three datasets. An exclusive review of SAR datasets is detailed in Table I. The experimental setup is detailed in Table II. Heridal has 640x640, Heridal(S0) has 1280x1280, and Heridal(S1) has 640x640 image resolutions with data augmentation like brightness, saturation, and exposure applied which are described in Table II. Our method of attention-based prediction head for YOLOv7-tiny and TPH-YOLOv5 is also evaluated on all three SAR datasets.

Table 1: Setup (2 GeForce RTX2080Ti GPUs and Intel i7-4790K@4GHzx8 CPU)

Dataset	**train images**	**val images**	**test images**	**shape**	**classes**	**instances**
Heridal	738	204	98	640x640	1	multiple
Mobdrone	1137	120	301	1920x1012	1	1 or 2
SeaDroneSee	648	129	88	5456x3632	1	multiple
Heridal(S0)	743	209	116	4000x3000	1	multiple
Heridal(S1)	738	204	98	640x640	1	multiple

4.1 YOLOv7 on Heridal data - Results

For experimentation on Heridal data, using YOLOv7 and a pre-trained model which has weights of YOLOv7 already trained on 640x640 image resolution for 300 epochs with COCO[31] data are considered. Corresponding tests were conducted tuning the hyper-parameters such as batch-size, learning rate, but with the same architecture of YOLOv7. Qualitative results for all tests are showcased in fig.4. Input images are scaled down to 640x640 resolution. Experimentation leads with set parameters, using which the CNN was trained from epoch 1 with the transfer learning technique. Training for 500epochs and 950epochs on heridal data are carried out respectively and tested with an intersection-over-union of 0.65, and the evaluation metrics are plotted in fig.7,8.

4.2 YOLOv7 on Mobdrone and SeaDronesSee data - Results

For experimentation on both the datasets, in the initialization step, we used YOLOv7 and trained weights of COCO[31]-640x640 image resolution for 300 epochs as similar to Heridal experiment. Experimentation leads with set parameters, using which the CNN was trained from epoch 1 with the transfer learning technique. Training for 300 epochs and 180 epochs respectively are carried out on all three datasets and tested with an intersection-over-union of 0.65, where other evaluation graphs are plotted. Qualitative results for tests are showcased in fig.5,6. The mAP, F1, Precision, and Recall for YOLOv7 when trained and tested for all three datasets are plotted and given in fig.7,8.

Fig. 4: Heridal APH-YOLOv7t detection (with persons in orange boxes)

5 Discussion

In this paper, we present the results of experiments conducted to test the performance of the state-of-the-art neural network based-solution YOLOv7 as mentioned in Sec.4, in the context of designing a vision system to perform search and rescue operations. The key problems like the pre-processing steps and data augmentation on drone data were addressed. Heridal data has smaller size person class entities whereby Heridal(S0) with 1280x1280 resolution on YOLOv7-tiny improved the mAP by +0.05. Heridal data(S1) when data augmentation is applied, that is saturation, brightness, and exposure of -25 to +25 percent has dropped mAP by -0.14 to that of Heridal data on YOLOv7-tiny; proving any change in the image parameters will affect the mAP drastically. APH-YOLOv7t has inference time of under 20ms whereas TPH-YOLOv5 has an inference time of over 100ms. APH-YOLOv7t has lower FLOPS and Parameters compared to TPH-YOLOv5 but is slightly higher than YOLOv7-tiny. Smaller models, in our case YOLOv7-tiny with fewer parameters tend to have faster inference times compared to slightly larger model APH-YOLOv7t, which is expected. YOLOv7-tiny is slightly lighter than your approach and, for the Heridal dataset, it presents higher results, although they are not optimal. Additionally, for the Mobdrones dataset, the performance is also higher than set 0.80 mAP. The reason why the inference time is found to be better for some models compared to others is because of model size, GPU, batch size, and model quantization.

Table 2: YOLO - Train and Test Results with Heridal data

Method	**Dataset**	**Size**	**Epochs**	**F1-Score**	**mAP50**	**Inference**
YOLOv7	Heridal	640x640	950	0.806	0.816	7.8ms
YOLOv7	Mobdrone	896x544	300	0.901	0.885	7.4ms
YOLOv7	SeaDronesSee	960x768	180	0.899	0.898	6.5ms
YOLOv7-tiny	Heridal	1280x1280	400	0.611	0.569	1.6ms
YOLOv7-tiny	Heridal(S0)	1280x1280	300	0.661	0.610	4.1ms
YOLOv7-tiny	Heridal(S1)	1280x1280	750	0.472	0.421	1.5ms
YOLOv7-tiny	Mobdrone	1280x1280	100	0.952	0.933	1.6ms
YOLOv7-tiny	SeaDronesSee	1280x1280	120	0.777	0.729	1.6ms
TPH-YOLOv5	Heridal	1280x1280	30	0.619	0.614	223ms
TPH-YOLOv5	Mobdrone	1280x1280	10	0.859	0.874	116ms
TPH-YOLOv5	SeaDronesSee	1280x1280	10	0.790	0.731	218ms
APH-YOLOv7t	Heridal	1280x1280	400	0.410	0.278	7.9ms
APH-YOLOv7t	Mobdrone	1280x1280	100	0.891	0.808	16.8ms
APH-YOLOv7t	SeaDronesSee	1280x1280	120	0.766	0.709	10.0ms

Table 3: YOLO - Computational complexity

Method	Layers	GFLOPS	Parameters
YOLOv7	314	103.2	36.4M
YOLOv7-tiny	208	13	6M
TPH-YOLOv5	371	160	41M
APH-YOLOv7t	300	13.5	7.2M

6 Conclusion

In this paper, the demonstrated APH-Yolov7t has delivered a competitive performance compared to the YOLOv7-tiny. Results show that, overall, YOLOv7 was more robust on the Heridal, Mobdrone, and SeaDronesSee datasets and comprehensive for our application. Our method APH-YOLOv7t, with attention mechanism on baseline, resulted in reasonably well in inference when compared to TPH-YOLOv5 and competitively well in mAP50 when compared to YOLOv7 and YOLOv7-tiny. Drone images in Search and Rescue Operations often have varying resolutions and scales, an attention-based head can improve the model's efficiency, potentially leading to better person detection. Our method supports such missions. YOLOv7-tiny is already a lightweight model known for its real-time performance. By optimizing the attention-based head for efficiency, one can maintain fast inference times while benefiting from the attention's contextual capabilities. For this, also there is a further chance of tuning TPH-YOLOv7t to transformer-based YOLOv7-tiny's head, which is expected to beat the baseline in terms of mAP50.

This widens up new possibilities for more widespread adoption of hybrid transformer prediction head-based YOLO models in the field, improving the efficiency of object detection models and reducing search time associated with

Fig. 5: Mobdrone APH-YOLOv7t detection (with persons in green boxes)

Fig. 6: SeaDronesSee APH-YOLOv7t detection (with persons in blue boxes)

drone deployment. Our study lays the foundation for future research and development in drones for search and rescue operations, ultimately leading to faster, more accurate and efficient practices in the field.

References

1. Murphy, R.; Griffin, C.; Stover, S.; Pratt, K. Use of micro air vehicles at hurricane katrina. In: IEEE Workshop on Safety Security Rescue Robots (2006)
2. Murphy, R.R., Disaster Robotics. MIT press, (2014)
3. "shannon l (2019) dji drones helped track and stop the notre dame fire. the verge". Available online: https://www.theverge.com/2019/4/16/18410723/notre-dame-fire-dji-drones-tracking-stopped-thermal-cameras.
4. Woo, S., Park, J.; Lee, J.-Y.; Kweon, I.S. Cbam: Convolutional block attention module. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 3–19 (2018).
5. Zhu, X.; Lyu, S.; Wang, X.; Zhao, Q. Tph-yolov5: Improved yolov5 based on transformer prediction head for object detection on drone-captured scenarios. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 2778–2788 (2021).
6. Božić-Štulić, D.; Marušić, Ž.; Gotovac, S. Deep learning approach in aerial imagery for supporting land search and rescue mission. *IJCV* 2019, *10*, 1256–1278.

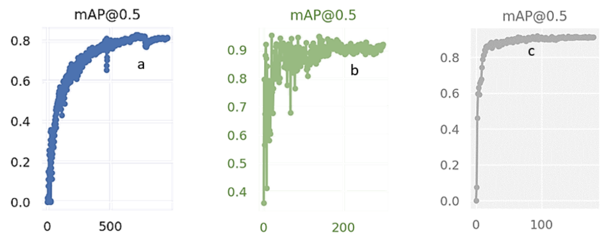

Fig. 7: Mean Average Precision (mAP) vs no. of epochs - plots for (a)Heridal, (b)Mobdrone, (c)SeaDronesSee

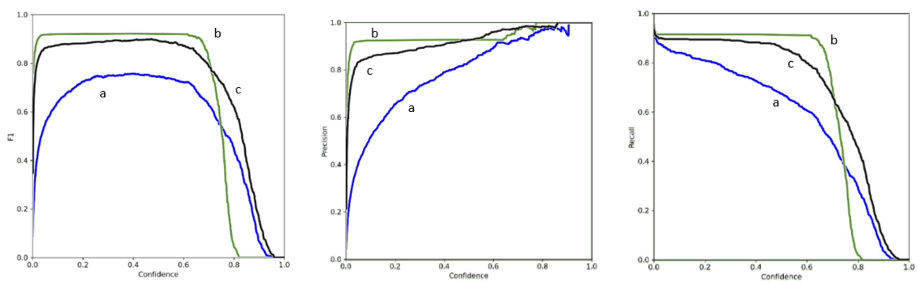

Fig. 8: F1, Precision and Recall graphs for (a)Heridal, (b)Mobdrone, (c)SeaDronesSee

7. Cafarelli, D.; Ciampi, L.; Vadicamo, L.; Gennaro, C.; Berton, A.; Paterni, M.; Benvenuti, C.; Passera, M.; Falchi, F. Mobdrone: A drone video dataset for man overboard rescue. In: Image Analysis and Processing–ICIAP 2022: 21st International Conference, Italy (2022). Springer.
8. Varga, L.A.; Kiefer, B.; Messmer, M.; Zell, A. Seadronessee: A maritime benchmark for detecting humans in open water. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (2022).
9. Zhao, J.-X.; Liu, J.-J.; Fan, D.-P.; Cao, Y.; Yang, J.; Cheng, M.-M. Egnet: Edge guidance network for salient object detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 8779–8788 (2019).
10. Qin, X.; Zhang, Z.; Huang, C.; Gao, C.; Dehghan, M.; Jagersand, M. Basnet: Boundary-aware salient object detection. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7471–7481 (2019)
11. Wang, W.; Lai, Q.; Fu, H.; Shen, J.; Ling, H.; Yang, R. Salient object detection in the deep learning era: An in-depth survey. *IEEE PAMI* 2008, 44(6), 3239-3259.
12. Kousik, N.; Natarajan, Y.; Raja, R.A.; Kallam, S.; Patan, R.; Gandomi, A.H. Improved salient object detection using hybrid convolution recurrent neural network. Expert Systems with Applications 166, 114064 (2021)
13. Hearst, M.A.; Dumais, S.T.; Osuna, E.; Platt, J.; Scholkopf, B. Support vector machines. IEEE Intelligent Systems and their applications 13(4), 18–28 (1998).
14. Girshick, R. Fast r-cnn. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1440–1448 (2015).

15. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems 28 (2015).
16. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767 (2018).
17. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. Ssd: Single shot multibox detector. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam (2016). Springer.
18. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollar, P. Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2980–2988 (2017).
19. Gotovac, S.; Zelenika, D.; Marušić, Ž.; Božić-Štulić, D. Visual-based person detection for search-and-rescue with uas: Humans vs. machine learning algorithm. *Remote Sensing* 12(20), 3295 (2020).
20. Ciccone, F.; Bacciaglia, A.; Ceruti, A. Methodology for image analysis in airborne search and rescue operations. In: Advances on Mechanics, Design Engineering IV: Proceedings of the International Joint Conference on Mechanics, Design Engineering & Advanced Manufacturing, JCM 2022, Italy Springer.
21. Caputo, S.; Castellano, G.; Greco, F.; Mencar, C.; Petti, N.; Vessio, G. Human detection in drone images using yolo for search-and-rescue operations. In: AIxIA 2021–Advances in Artificial Intelligence: 20th International Conference of the Italian Association for Artificial Intelligence, Virtual Event, 2021. Springer.
22. Gordienko, Y.; Rokovyi, O.; Alienin, O.; Stirenko, S. Context-aware data augmentation for efficient object detection by uav surv. In: 2022 IEEE 10th International Symposium on Digital Forensics and Security (ISDFS) (2022).
23. Rahnemoonfar, M.; Chowdhury, T.; Sarkar, A.; Varshney, D.; Yari, M.; Murphy, R.R. Floodnet: A high resolution aerial imagery dataset for post flood scene understanding. *IEEE Access* (2021).
24. Poddar, N.; Jain, S. Light weight character and shape recognition for autonomous drones. arXiv preprint arXiv:2208.06804 (2022).
25. Sambolek, S.; Ivasic-Kos, M. SEARCH AND RESCUE IMAGE DATASET FOR PERSON DETECTION - SARD, https://doi.org/10.21227/ahxm-k331
26. Lacmus data. https://github.com/lacmus-foundation/lacmus.
27. Wang, C.-Y.; Bochkovskiy, A.; Liao, H.-Y.M. Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2023).
28. Dousai, N.M.K.; Loncaric, S. Detection of humans in drone images for search and rescue operations. In: 2021 3rd Asia Pacific Information Technology Conference, pp. 69–75 (2021)
29. Patrik, A., Utama, G., Gunawan, A.A.S. et al. GNSS-based navigation systems of autonomous drone for delivering items. J Big Data 6, 53 (2019). https://doi.org/10.1186/s40537-019-0214-3
30. Roboflow. https://roboflow.com/
31. Coco dataset. https://cocodataset.org/
32. Kaufmann, Elia, Antonio Loquercio, Rene Ranftl, Alexey Dosovitskiy, Vladlen Koltun, and Davide Scaramuzza. "Deep drone racing: Learning agile flight in dynamic environments." In Conference on Robot Learning, pp. 133-145. PMLR, 2018.
33. Valenti, Francesco, Domenico Giaquinto, Luigi Musto, Andrea Zinelli, Massimo Bertozzi, and Alberto Broggi. "Enabling computer vision-based autonomous navigation for unmanned aerial vehicles in cluttered gps-denied environments." In 2018

21st International Conference on Intelligent Transportation Systems (ITSC), pp. 3886-3891. IEEE, 2018.

34. Tomic, Teodor, Korbinian Schmid, Philipp Lutz, Andreas Domel, Michael Kassecker, Elmar Mair, Iris Lynne Grixa, Felix Ruess, Michael Suppa, and Darius Burschka. "Toward a fully autonomous UAV: Research platform for indoor and outdoor urban search and rescue." IEEE robotics automation magazine 19, no. 3 (2012): 46-56.
35. Schilling, Fabian, Fabrizio Schiano, and Dario Floreano. "Vision-based drone flocking in outdoor environments." IEEE Robotics and Automation Letters 6, no. 2 (2021): 2954-2961.
36. Lyu, Mingyang, Yibo Zhao, Chao Huang, and Hailong Huang. 2023. "Unmanned Aerial Vehicles for Search and Rescue: A Survey" Remote Sensing 15, no. 13: 3266. https://doi.org/10.3390/rs15133266