Open Problems in Computer Vision for Wilderness SAR and The Search for Patricia Wu-Murad

Thomas Manzini, Dr. Robin Murphy
Texas A&M University
400 Bizzell St, College Station, TX 77843
tmanzini@tamu.edu, robin.r.murphy@tamu.edu

Abstract

This paper details the challenges in applying two computer vision systems, an EfficientDET supervised learning model and the unsupervised RX spectral classifier, to 98.9 GB of drone imagery from the Wu-Murad wilderness search and rescue (WSAR) effort in Japan and identifies 3 directions for future research. There have been at least 19 proposed approaches and 3 datasets aimed at locating missing persons in drone imagery, but only 3 approaches (2 unsupervised and 1 of an unknown structure) are referenced in the literature as having been used in an actual WSAR operation. Of these proposed approaches, the EfficientDET architecture and the unsupervised spectral RX classifier were selected as the most appropriate for this setting. The EfficientDET model was applied to the HERIDAL dataset and despite achieving performance that is statistically equivalent to the state-of-the-art, the model fails to translate to the real world in terms of false positives (e.g., identifying tree limbs and rocks as people), and false negatives (e.g., failing to identify members of the search team). The poor results in practice for algorithms that showed good results on datasets suggest 3 areas of future research: more realistic datasets for wilderness SAR, computer vision models that are capable of seamlessly handling the variety of imagery that can be collected during actual WSAR operations, and better alignment on performance measures.

1. Introduction

Patricia Wu-Murad is a 60-year-old female hiker who went missing along the Komado Komo trail in southern Nara, Japan on April 10, 2023. The search for her included hundreds of wilderness search and rescue (WSAR) personnel, several of which operated small unmanned aerial systems (sUAS) that collected 98.9 GB of imagery. This volume of imagery can be difficult to inspect manually, so these operations were augmented by using CV techniques
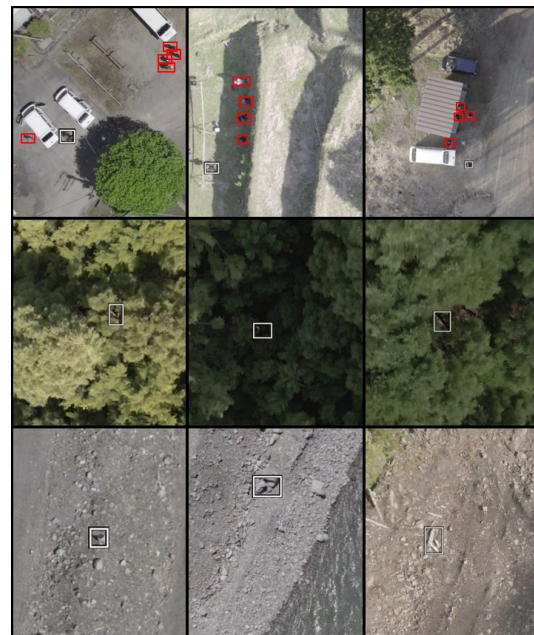

Figure 1. Predictive failures from *Tiled_EfficientDET*$_{174}$. Row 1: Failure to find members of the search team (false negatives in red). Row 2: Misclassification of tree limbs as people. Row 3: Misclassification of rocks as people. Errors corresponding to rows 2 and 3 were particularly common in the data collected in the field.

to identify specific regions of interest that may be used to help focus search efforts. This paper details the efforts and challenges in developing and deploying CV systems in an actual WSAR operation and highlights 3 future directions for research. At the time of writing Patricia Wu-Murad remains missing[1].

2. Wilderness SAR Operations

WSAR is a class of field operations that involves searching for and rescuing a missing person in austere wilderness

---

[1]Imagery, code, models, and instructions for replication can be found at https://github.com/CRASAR/WiSAR/

settings. This is done by sending teams of people to the areas where the missing person is suspected to be located where they will perform search activities such as ground search patterns (e.g., grid-search), following scent trails using dogs, or flying drones to collect aerial imagery to uncover information that will lead to the missing person's location.

2.1. sUAS for Wilderness SAR

sUAS are small, portable aerial robots with onboard sensors and cameras that can be controlled by a ground crew to search an area for the missing person or signs of their presence [29]. In WSAR settings, these sUAS have become valuable as they enable ground teams to collect high resolution imagery over a large area. These sUAS have been widely used in WSAR [30, 19] and have resulted in the rescue of hundreds of missing persons [1].

sUAS operations for WSAR can take several forms. For example, hasty searches are flown at several hundred feet above the ground in a boustrophedon or lawnmower pattern collecting nadir imagery at fixed intervals. In some situations, flight plans can be designed that attempt to maximize the probability of detection of a person based on lost person behavior models [50]. Other approaches consider flight paths for fixed wing vehicles that allow the collection of multiple views of objects without the ability to hover [6].

Even though the operators of the sUAS are typically able to inspect the aerial view while in flight, they are usually capturing high definition imagery that can be further analyzed, a process colloquially referred to as "squinting." It is common for search teams to spend hours observing the imagery in an effort to ensure that no clues or regions of interest were missed.

2.2. Computer Vision for Wilderness SAR

In supporting WSAR operations, the efforts of the CV community have been primarily focused on automating the squinting process described in Section 2.1.

On the unsupervised side, several color-based approaches have been proposed [31, 27, 33], and one histogram-based approach was discussed in [4].

On the supervised learning side, three datasets dominate the literature: the SARD dataset [42], the HERIDAL dataset [28], and most recently, the WiSARD dataset, which explores both electro-optical and infrared modalities [12]. These three datasets focus on person detection in wilderness settings and are annotated using bounding boxes.

At least 19 models have been trained on these datasets, and all of them lean heavily on the existing literature in object detection. The YOLO architecture [38] is by far the most prolific in this space as multiple variants have been trained and evaluated on the SARD dataset [42, 9, 14, 5, 20], the HERIDAL dataset [14, 34], and the WiSARD dataset
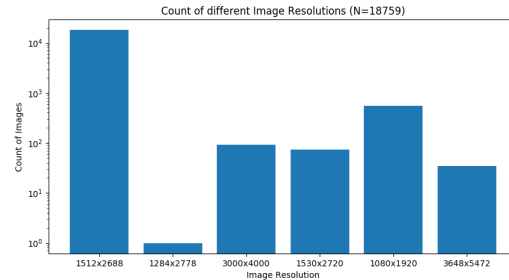

Figure 2. The count of each distinct image resolution in the sampled frames collected during the search for Patricia Wu-Murad.

[12]. The Single Shot Detection architecture [25] and the FasterRCNN [39] architectures have been explored in [23] and [28, 41], respectively. The RetinaNet architecture [24] was also extensively experimented with in [37, 36]. The EfficientDET architecture was explored in [16]; however, they did so on an unpublished subset of cropped HERIDAL samples, making it difficult to compare its effectiveness.

Outside of approaches driven by these datasets, other authors explore supervised CV systems to find missing persons in avalanches [7] and search for missing persons in maritime settings[40, 26], along side other applications [8].

Based on a review of the literature, at least two unsupervised methods have been used in a real world search and rescue operation: The RX spectral classifier described in [31, 27] was used in the Texas Memorial Day Floods in 2015 [35], and the commercial product Loc8 [2, 10], a system which relies on color based methods, has also seen operational use [45]. The commercial product SARUAV [3] has also been used operationally and claims to use a deep neural network, though it does not provide details on training data or model architecture [32].

3. Drone Operations & Data Details

The search for Patricia Wu-Murad, generated 98.9 GB of imagery over approximately 3 weeks. This data consisted of 128 high resolution photos and 243 high resolution videos at different resolutions as shown in Figure 2.

Teams primarily flew two types of missions: high altitude nadir imagery and low altitude oblique imagery. High altitude nadir imagery is the standard in WSAR operations; however, as tree cover in this region of Japan is dense, drone teams began collecting low altitude oblique imagery below the treeline. Thus high altitude nadir imagery was reserved for open areas such as river beds, farmland, and trails.

In order to process this data in a reasonable amount of time, all videos were sampled at a rate of 2 frames per second, and all images were processed as is. This generated a set of 18,759 images to be analyzed. A sample of these photos can be seen in Figure 3.

Figure 3. Sample images selected from the 18,759 frames analyzed during the search for Patricia Wu-Murad.

It is worth noting that because these flights were conducted over a mountainous region and at varying altitudes and resolutions there is substantial scale variance in potential regions of interest and pixels per inch.

4. Computer Vision Techniques Employed

As discussed in Section 3, the imagery consisted of both oblique and nadir views. Efforts were directed to the high altitude, mostly nadir imagery, because the low altitude oblique imagery was collected such that any potential region of interest be immediately evident, and the high altitude imagery is the most time intensive to review.

Both a supervised and unsupervised method were employed to support the search efforts: an augmented version of the RX spectral classifier detailed in [31, 35], and an EfficientDet bounding box model [47, 49] trained on the HERIDAL dataset. These two methods are discussed herein.

4.1. Unsupervised Approach

The RX classifier was employed because one of the ground WSAR teams had already utilized the Loc8 software on this imagery and because the RX classifier had already seen use in previous WSAR field operations. As discussed in [35], this method selects anomalous pixels in an image based on the mahalanobis distance function and a given p-value resulting in almost all images having some set of anomalous pixels. This would mean that the output of this classifier must be inspected for every image. However, manual inspection of the 18,759 images was not practical. As a result, additional post processing steps were performed in order to further refine the classifier outputs. DBSCAN clustering was performed on the detected anomalous pixels in order to group anomalies together. Photos that had <1 and >4 clusters and clusters that contained <209 pixels were discarded.

This approach resulted in 5234 candidate images that were all manually reviewed. Following that review, none of the candidate images were elevated to the search team.[2]

---

[2]Hyperparameters and sample outputs are included in Appendix B.

4.2. Supervised Approach

This section details the building and the performance of a supervised model to support the ongoing WSAR activities.

4.2.1 Choice of Dataset

The 3 principal datasets referenced in the literature were considered: SARD, HERIDAL, and WiSARD. SARD was discarded as its altitude, view angles, and target class poses were too different from the data that had been collected in the field and the inspection tasks required in this effort. The WiSARD dataset was also not pursued as it has only recently been released and has not been evaluated except by its publishing authors. Instead, the HERIDAL dataset was utilized as it partially aligned with the imagery inspection tasks at hand and it had been studied academically.

4.2.2 Choice of Model Architecture

As detailed in Section 2.2, the most explored network architectures in this space are YOLO and RetinaNet. The RetinaNet-based AIR model architecture detailed in [37, 36] was considered, but not pursued, due to concerns about replicability stemming from the sensitivity to initial network conditions referenced in [36]. Following this, the YOLO [21] and EfficientDET [46] architectures were explored and training on the HERIDAL dataset began.

Given the active WSAR operations, the YOLOv5 approach was abandoned because poor validation set performance was observed compared to the EfficientDET model.

The model pursued was a 512x512 EfficientDET model and training samples were generated by selecting appropriately sized crops of from the HERIDAL dataset.[3] Inference was then performed by regularly spacing 512x512 tiles from high resolution imagery as described in [36]. All predicted bounding boxes with overlap are merged into a new bounding box surrounding the overlapping boxes.

Two versions of this model are reported here: $Tiled\_EfficientDET_{84}$ was run as a hasty effort to collect results to share with search teams, and $Tiled\_EfficientDET_{174}$ was run following further training. The number that follows the model name denotes the number of epochs that the model was trained for. Both models were run on the imagery collected in the field.

Following the termination of search efforts, these models were evaluated on the HERIDAL test set and the results are reported in Table 1. $Tiled\_EffecientDET_{174}$ receives an average precision score of 91.5, which is statistically equivalent to the current SOTA model performance ($p \leq 0.02$), without tuning hyperparameters on the test set as described in [36].

---

[3]The details of this model training process are included in Appendix C.

	P	R	AP
VOC2012 Evaluation [17]			
Mean Shift Segmentation [48, 36]	18.7	74.7	-
Saliency Guided VGG16 [11]	34.8	88.9	-
Faster R-CNN (2019) [11]	58.1	85	-
Faster R-CNN (2018) [28]	67.3	88.3	**86.1**
Two-stage Multimodel CNN [23]	68.9	94.7	-
SSD [23]	4.3	94.4	-
YOLOv5s [14]	75.3	69.4	73.1
YOLOv5m [14]	79.7	81.2	81
AIR with NMS** [36]	90.1	86.1	84.6
Tiled_EffecientDET$_{84}$ (ours)	58.0	53.7	49.3
Tiled_EffecientDET$_{174}$ (ours)	78.6	84.0	80.1
SAR-APD Evaluation [36]			
AIR with NMS** [36]	90.5	87.8	86.5
AIR with MOB** [36]	94.9	92.9	**91.7**
Tiled_EffecientDET$_{84}$ (ours)	65.3	61.4	61.3
Tiled_EffecientDET$_{174}$ (ours)	82.8	91.4	**91.5**

Table 1. Precision (P), Recall (R), and Average Precision (AP) on the HERIDAL test set. Entries marked in bold are either SOTA performance or are within the 95% confidence interval. Precision and Recall scores are not bolded because of inconsistencies in the literature regarding thresholds. Precision and Recall scores reported for *Tiled_EffecientDET* models are reported with a 0.5 threshold. Entries marked with ** were attained following hyper-parameter tuning on the test set.

4.2.3 Inference Details & Model Failures

Inference of the EfficientDET models on the 18,759 images was performed. *Tiled_EffecientDET*$_{84}$ generated bounding boxes in 1522 images at 50% confidence, and *Tiled_EffecientDET*$_{174}$ generated bounding boxes in 1377 images at 60% confidence, all of which were manually reviewed. This process resulted in 4 candidate images that were elevated to the search teams[4].

Unfortunately, these models generate a substantial number of false positives and negatives, routinely failing to recognize search team members and frequently mislabeling bushes, rocks, and branches as people as shown in Figure 1. Finally, chromatic aberrations near photo edges are occasionally labeled as people, presumably because of red and green coloration.

5. Future Directions

The current body of academic literature leaves much to be desired when it comes to CV applications for WSAR. There are three directions that are needed to better refine this area of work: more realistic and diverse datasets, models which leverage the data that is available in practice, and community alignment on evaluation metrics.

---

[4]These 4 candidate images are included in Appendix D.

More Realistic Datasets. The three prominent datasets in this space (SARD, HERIDAL, and WiSARD) frequently contain samples that are not realistic to WSAR. Targets in these datasets often wear bright, clean clothing, despite missing persons often being covered in mud and dirt when found [22]. Targets are frequently seen holding drone controllers, something that missing persons would likely not possess. Targets are also frequently located in open areas or on trails, despite missing persons typically seeking shelter and descending to lower terrain [22]. Finally, objects such as backpacks, clothing, and other evidence of human activity are frequently excluded from datasets, or worse, included intentionally as false positive samples despite being valuable to search efforts. All of these factors will induce model biases that may result in failures to find missing persons when deployed at scale.

Better Models & Features. The majority of models in the literature are straightforward instantiations of popular CV models that have been trained solely on WSAR imagery. However, in practice, there are substantial amounts of information that can be used to augment these models. For example, current models need to manage the scale variance problem described in Section 3 through data augmentation. But one could imagine learning deformable convolutions [15] conditioned on image resolution, focal length, and flight altitude (similar to the analysis done in [18]) to make the scale variances an easier function to learn. Other models could be trained to leverage details of the missing person, such as age or clothing color. Features like these need to be considered when constructing datasets so future models can leverage the data present during real world operations.

Alignment on Evaluation Metrics. There are at least five papers that do not report average precision scores for their models. While, all papers report precision and recall, or statistics that can be used to compute those measures, authors rarely report their model's decision thresholds. Inconsistent reporting of performance metrics makes it difficult to understand how different approaches compare to one another. As argued in [36], this context strongly suggests that the SAR-APD metric is a more appropriate evaluation scheme than VOC2012 [17]. Thus, future papers are encouraged to report Average Precision based on the SAR-APD bounding box evaluation scheme as reported here and in [36].

6. Conclusion

This paper discussed the current state of the art in using sUAS and CV for WSAR; detailed the operational environment, the data collection and the deployment of CV methods on an actual missing persons case; and highlighted three future directions for the research community.

References

[1] DJI Drone Rescue Map. https://enterprise.dji.com/drone-rescue-map. Accessed: 2023-06-15.

[2] Loc8 software. https://help.loc8.life/. Accessed: 2023-06-15.

[3] SARUAV software. https://www.saruav.pl/. Accessed: 2023-06-15.

[4] M Talha Agcayazi, Eric Cawi, Arsenie Jurgenson, Parham Ghassemi, and Gerald Cook. Resquad: Toward a semi-autonomous wilderness search and rescue unmanned aerial system. In *2016 International Conference on Unmanned Aircraft Systems (ICUAS)*, pages 898–904. IEEE, 2016.

[5] Namat Bachir and Qurban Memon. Investigating yolov5 for search and rescue operations involving uavs: Investigating yolo5. In *Proceedings of the 5th International Conference on Control and Computer Vision*, pages 200–204, 2022.

[6] Ashvin Bashyam and Jacob Guggenheim. Uavs for wilderness search and rescue: Real-world considerations and technology roadmap for fixed wing uavs. *Journal of Search and Rescue*, 3(1), 2019.

[7] Mesay Belete Bejiga, Abdallah Zeggada, Abdelhamid Nouffidj, and Farid Melgani. A convolutional neural network approach for assisting avalanche search and rescue operations with uav imagery. *Remote Sensing*, 9(2):100, 2017.

[8] Arturo Miguel Russell Bernal and Jane Cleland-Huang. Hierarchically organized computer vision in support of multi-faceted search for missing persons. In *2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG)*, pages 1–7. IEEE, 2023.

[9] Alessandro Betti and Mauro Tucci. Yolo-s: A lightweight and accurate yolo-like network for small target detection in aerial imagery. *Sensors*, 23(4):1865, 2023.

[10] Ryan Blackburn and Robert Joslin. Detectability of clothing color by small unmanned aircraft systems for search and rescue operations. *Journal of Aviation/Aerospace Education & Research*, 30(1):41–88, 2021.

[11] Dunja Božić-Štulić, Željko Marušić, and Sven Gotovac. Deep learning approach in aerial imagery for supporting land search and rescue missions. *International Journal of Computer Vision*, 127(9):1256–1278, 2019.

[12] Daniel Broyles, Christopher R Hayner, and Karen Leung. Wisard: A labeled visual and thermal image dataset for wilderness search and rescue. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 9467–9474. IEEE, 2022.

[13] A. Buslaev, A. Parinov, E. Khvedchenya, V. I. Iglovikov, and A. A. Kalinin. Albumentations: fast and flexible image augmentations. *ArXiv e-prints*, 2018.

[14] Sergio Caputo, Giovanna Castellano, Francesco Greco, Corrado Mencar, Niccolò Petti, and Gennaro Vessio. Human detection in drone images using yolo for search-and-rescue operations. In *International Conference of the Italian Association for Artificial Intelligence*, pages 326–337. Springer, 2021.

[15] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. Deformable convolutional networks. In *Proceedings of the IEEE international conference on computer vision*, pages 764–773, 2017.

[16] Nayee Muddin Khan Dousai and Sven Lončarić. Detecting humans in search and rescue operations based on ensemble learning. *IEEE Access*, 10:26481–26492, 2022.

[17] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results. http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html.

[18] Michael A Goodrich, Bryan S Morse, Damon Gerhardt, Joseph L Cooper, Morgan Quigley, Julie A Adams, and Curtis Humphrey. Supporting wilderness search and rescue using a camera-equipped mini uav. *Journal of Field Robotics*, 25(1-2):89–110, 2008.

[19] Sean Grogan, Robert Pellerin, and Michel Gamache. The use of unmanned aerial vehicles and drones in search and rescue operations–a survey. *Proceedings of the PROLOG*, pages 1–13, 2018.

[20] Seungoh Han, Ah-Young Nho, Wei Teng Kwan, Benjamin Paglia, Jacob Visniski, Minji Lee, Eric T Matson, and Minsun Lee. Ehdnet: Enhanced human detection network for search and rescue. In *2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC)*, pages 448–449. IEEE, 2022.

[21] Glenn Jocher, Ayush Chaurasia, Alex Stoken, Jirka Borovec, Yonghye Kwon, Kalen Michael, Jiacong Fang, Zeng Yifu, Colin Wong, Diego Montes, et al. ultralytics/yolov5: V7. 0-yolov5 sota realtime instance segmentation. *Zenodo*, 2022.

[22] R.J. Koester. *Lost Person Behavior: A Search and Rescue Guide on where to Look for Land, Air, and Water*. dbS Productions, 2008.

[23] Mirela Kundid Vasić and Vladan Papić. Multimodel deep learning for person detection in aerial images. *Electronics*, 9(9):1459, 2020.

[24] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017.

[25] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, pages 21–37. Springer, 2016.

[26] Eleftherios Lygouras, Nicholas Santavas, Anastasios Taitzoglou, Konstantinos Tarchanidis, Athanasios Mitropoulos, and Antonios Gasteratos. Unsupervised human detection with an embedded vision system on a fully autonomous uav for search and rescue operations. *Sensors*, 19(16):3542, 2019.

[27] Travis Marshall and L Nathan Perkins. Color outlier detection for search and rescue. *Dept. Elect. Comput. Eng., Boston Univ., Boston, MA, USA, Tech. Rep. ECE-2015-01*, 2015.

[28] Željko Marušić, Dunja Božić-Štulić, Sven Gotovac, and Tonćo Marušić. Region proposal approach for human detection on aerial imagery. In *2018 3rd International Conference on Smart and Sustainable Technologies (SpliTech)*, pages 1–6. IEEE, 2018.

[29] Sven Mayer, Lars Lischke, and Paweł W Woźniak. Drones for search and rescue. In *1st International Workshop on Human-Drone Interaction*, 2019.

[30] Jake N McRae, Christopher J Gay, Brandon M Nielsen, and Andrew P Hunt. Using an unmanned aircraft system (drone) to conduct a complex high altitude search and rescue operation: a case study. *Wilderness & environmental medicine*, 30(3):287–290, 2019.

[31] B. Morse, Daniel R. Thornton, and Michael A. Goodrich. Color anomaly detection and suggestion for wilderness search and rescue. *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 455–462, 2012.

[32] Tomasz Niedzielski, Mirosława Jurecka, Bartłomiej Miziński, Wojciech Pawul, and Tomasz Motyl. First successful rescue of a lost person using the human detection system: A case study from beskid niski (se poland). *Remote Sensing*, 13(23):4903, 2021.

[33] Tomasz Niedzielski, Mirosława Jurecka, Magdalena Stec, Małgorzata Wieczorek, and Bartłomiej Miziński. The nested k-means method: A new approach for detecting lost persons in aerial images acquired by unmanned aerial vehicles. *Journal of Field Robotics*, 34(8):1395–1406, 2017.

[34] Alexandru Gabriel Popa, Loretta Ichim, and Dan Popescu. Real-time person detection from uav images using performant neural networks. In *2022 14th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, pages 1–5. IEEE, 2022.

[35] Julia Proft, Jesus Suarez, and Robin Murphy. Spectral anomaly detection with machine learning for wilderness search and rescue. In *2015 IEEE MIT Undergraduate Research Technology Conference (URTC)*, pages 1–3. IEEE, 2015.

[36] Pasi Pyrrö et al. Air: Aerial inspection retinanet for land search and rescue missions. 2022.

[37] Pasi Pyrrö, Hassan Naseri, and Alexander Jung. Rethinking drone-based search and rescue with aerial person detection. *arXiv preprint arXiv:2111.09406*, 2021.

[38] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

[39] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 2015.

[40] MP Sahana, Aniket Sengar, Aniruddh Dubey, and Umang Agrawal. Person detection in maritime search and rescue operations. 2022.

[41] Sasa Sambolek and Marina Ivasic-Kos. Person detection in drone imagery. In *2020 5th International Conference on Smart and Sustainable Technologies (SpliTech)*, pages 1–6. IEEE, 2020.

[42] Sasa Sambolek and Marina Ivasic-Kos. Automatic person detection in search and rescue operations using deep cnn detectors. *Ieee Access*, 9:37905–37922, 2021.

[43] Ujjwal Saxena. Automold–road-augmentation-library. https://github.com/UjjwalSaxena/Automold–Road-Augmentation-Library/.

[44] Roman Solovyev, Weimin Wang, and Tatiana Gabruseva. Weighted boxes fusion: Ensembling boxes from different object detection models. *Image and Vision Computing*, 107:104117, 2021.

[45] TBnewsWatch.com Staff. More aircraft join search for missing cessna on 5th day. https://www.snnewswatch.com/local-news/more-aircraft-join-search-for-missing-cessna-on-5th-day-6648365.

[46] Mingxing Tan and Quoc Le. Efficientnetv2: Smaller models and faster training. In *International conference on machine learning*, pages 10096–10106. PMLR, 2021.

[47] Mingxing Tan, Ruoming Pang, and Quoc V Le. Efficientdet: Scalable and efficient object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10781–10790, 2020.

[48] Hrvoje Turić, Hrvoje Dujmić, and Vladan Papić. Two-stage segmentation of aerial images for search and rescue. *Information Technology and Control*, 39(2), 2010.

[49] Ross Wightman. Python effdet package. https://pypi.org/project/effdet/.

[50] R Williams. Collaborative multi-robot multi-human teams in search and rescue. In *Proceedings of the International ISCRAM Conference*, volume 17, 2020.