ACCEPTED VERSION
2023 IEEE International Symposium on Safety, Security and Rescue Robotics.
DOI Pending. Please find the published version on http://ieeexplore.ieee.org.

Disaster area recognition from aerial images with complex-shape class detection

Rubén González-Navarro, Dahui Lin-Yang, Ricardo Vázquez-Martín, and Alfonso Garcia-Cerezo

***Abstract*— This paper presents a convolutional neural network (CNN) model for event detection from Unmanned Aerial Vehicles (UAV) in disaster environments. The model leverages the YOLOv5 network, specifically adapted for aerial images and optimized for detecting Search and Rescue (SAR) related classes for disaster area recognition. These SAR-related classes are person, vehicle, debris, fire, smoke, and flooded areas. Among these, the latter four classes lead to unique challenges due to their lack of discernible edges and/or shapes in aerial imagery, making their accurate detection and performance evaluation metrics particularly intricate. The methodology for the model training involves the adaptation of the pre-trained model for aerial images and its subsequent optimization for SAR scenarios. These stages have been conducted using public datasets, with the required image labeling in the case of SAR-related classes. An analysis of the obtained results demonstrates the model's performance while discussing the intricacies related to complex-shape classes. The model and the SAR datasets are publicly available.**

Keywords: Deep Learning; Convolutional Neural Network; Transfer Learning; Aerial Images; Disaster Robotics

I. Introduction

Natural disasters have always been uncontrollable and unpredictable forces that have inflicted human suffering. The devastating impact of events such as avalanches or fires results in numerous casualties and poses immense challenges for Search and Rescue (SAR) operations. In the last decade, robotics solutions have emerged as invaluable tools to reduce the time employed on victim recovery and assistance. Applying robotics systems in SAR missions can improve the performance of tasks or reduce the risk to emergency staff. Many different works can be found applied in SAR issues: exploring the environment [1], victim monitoring and triage [2], communication formation [3] and victims recovery [4], [5].

The recreation of disasters scenarios has led to the publication of datasets related to SAR operations. Many datasets can be found for research purposes, such as AIDER [6] which contains a variety of disaster aerial images for image retrieval [7]. There are also multimodal SAR datasets like [8] which use thermal-RGB data to detect people. Recent popular datasets are related to semantic segmentation purposes, where RescueNet [9] is a high-resolution UAV semantic segmentation dataset to provide a benchmark for image segmentation in disaster areas. It classifies buildings' damages and segments different elements such as debris, water, or pools. Nevertheless, there are not many datasets published for segmentation tasks.

Most existing SAR datasets focus on detecting victims or classifying disaster areas, which significantly aids victim rescue efforts. UAVs play a crucial role in these tasks due to their versatility and mobility across different scenarios, involving aerial images as an important data source. For instance, efforts to prevent wildfires have employed Convolutional Neural Networks (CNNs) to detect potential signs of forest fires using aerial RGB images [10], [11]. Similarly, for assessing building damages, authors [12]–[14] propose a CNN to detect roofs from UAV images and evaluate the extent of the damage. Also, CNN is used to search people with only RGB images, as it is done in [15], and by incorporating thermal infrared information for people detection [16], [17].

Despite the solutions proposed in previous works, they are focused on detecting a single class (e.g. fire, building, or people) limiting the scope of situational awareness. The main contribution of this paper is an aerial dataset for disaster scenarios with up to seven SAR-related classes corresponding to objects or events commonly encountered in disaster situations. In addition, a YOLOv5 model has been adapted to aerial imagery and optimized using our dataset for SAR missions. The model and the SAR dataset are publicly available [1].

The remaining of this paper is organized as follows: Section II outlines the methodology used in this work, Section III presents the datasets employed and the proposed SAR-related classes, Section IV summarizes and discusses the results and, finally Section V is devoted to the concluding remarks.

II. Methodology

In order to adapt the YOLOv5 network for aerial SAR images, two stages have been conducted: firstly adapting the pre-trained YOLOv5 network to aerial images, and finally, using transfer learning to optimize the model for disaster scenarios.

The YOLOv5 network, our chosen model for this work, is pre-trained for images at ground level, which typically exhibit significant differences in object appearance compared to aerial imagery. In the case of aerial images, objects tend to be much smaller and appear with lower resolution, which presents a challenge for accurate class detection. In addition,

*This work has been partially funded by the Spanish Ministerio de Ciencia, Innovación y Universidades, Gobierno de España, project PID2021-122944OB-I00.

The authors are with Robotics and Mechatronics Group, Institute of Mechatronics Engineering and Cyber-physical Systems, University of Malaga, Spain.

[1] https://github.com/ricardovmartin/TFGAerialDisaster

Fig. 1: Some sample images from the VisDrone dataset [18].

Fig. 2: Some images from the UMA-SAR dataset [20].

these models rely on metrics like the Intersection over Union (IoU), which is often very difficult to get high mean Average Precision (mAP) for 50 % IoU threshold. Network-generated bounding boxes and ground-truth (labels) often have to generate an IoU greater than 50 % in very few pixels. As a result of the aforementioned problems, a well-known public dataset in urban scenarios, namely VisDrone [18], has been used to train the model in aerial images. This dataset has been selected not only by the set of labeled images but also as a valuable benchmark that hosts a competition that encourages the development of network models capable of achieving high precision. Among these models, the YOLOv5 network is part of the evaluated and compared networks.

Once the model has been optimized for aerial images, the aim of training in disaster scenarios is to detect a series of SAR-related classes that are relevant to describe significant events in disaster areas. The previous stage of adapting the network to aerial views is essential to ensure that the model can effectively learn from aerial imagery before proceeding to the subsequent step of optimizing the model for aerial disaster images. In this stage, the adapted model for aerial images is optimized using transfer learning techniques to handle the new SAR-related classes using public SAR datasets. The UMA-SAR dataset [19] has been used, but no labels are available. To address data scarcity and ensure the model's effectiveness in detecting the desired classes, the AIDER dataset [6] has been used to have more examples of fire, smoke, and flooded areas.

III. Data model

III-A. Datasets

As described in the previous section, three different datasets have been used: VisDrone [18], UMA-SAR [19] and AIRDER [6].

The first dataset used in this work comprised labeled images in urban scenarios. These images were employed in the first stage, as explained in Section II, in order to adapt the model to aerial imagery. The VisDrone dataset has been curated and labeled by the AISKYEYE team of the Machine Learning and Data Mining Laboratory of Tianjin University, China [18]. This dataset is a valuable resource specifically designed to facilitate the adaptation of computer vision models to aerial views. It comprises a diverse collection of frames extracted from both video clips and still images, all of which were captured using several cameras mounted on UAVs. It should be noted that the dataset was collected under different scenarios and under different weather and lighting conditions in urban scenarios. These frames were manually annotated with over 2.6 million bounding boxes. Some important attributes, such as scene visibility, object class, and occlusion, are also provided for better use of the data. Some example images of this dataset are shown in Figure 1.

The other two datasets are used in the second stage for training in disaster scenarios. The first dataset used for optimizing the model in the SAR domain is the UMA-SAR dataset [19]. This dataset was collected during an annual Workshop organized by the Chair of Security, Emergencies, and Disasters at Universidad de Málaga (UMA), Spain. Some images can be seen in Figure 2. The images used for training were extracted from three videos captured from UAVs during the SAR exercises. To ensure a diverse and representative set of SAR scenes and to mitigate overfitting, we carefully selected frames from these videos. Although the UMA-SAR dataset is a valuable set of aerial images in a disaster area, it lacks ground-truth labels. So the set of 195 selected images was labeled, using the SAR-related classes as described in Section III.

However, in order to achieve a balanced dataset and to cover all relevant SAR-related classes, it was necessary to incorporate additional images from another source. Thus, we integrated images from the AIDER dataset, which consists of a collection of images of four types of disasters: fire/smoke, flood, building collapse/rubble, and traffic accidents, as well as non-disaster environments [6]. These aerial images are compiled from multiple sources, such as image search engines (Google and Bing images), YouTube, and news agency websites, as well as other aerial image databases. Figure 3 shows some examples of this dataset. The different catastrophe events in the dataset were captured with different resolutions and under different lighting and perspective conditions. The AIDER dataset was originally intended for image retrieval rather than detection, so we also labeled this dataset with SAR-related classes used in our disaster detection work.

By combining the UMA-SAR and AIDER datasets, we

Fig. 3: Example images of AIDER dataset [6].

created a comprehensive and balanced training dataset. This dataset enhances our model to effectively detect and classify SAR-related classes in aerial imagery, providing the capabilities for successful disaster scenario recognition.

III-B. Aerial SAR-representative classes

A total of 505 images were manually labeled for training. The labeled dataset consists of selected UMA-SAR and AIDER images distributed in three sets: training (70 % of total), validation (15 % of total), used during network training to adjust model parameters and avoid overfitting, and test (15 % of the total) used to test the model after training (inference).

We propose the next relevant set of objects and events related to SAR scenarios to train our model:

- **Fire**: this class is essential for identifying the presence of fire in a disaster area.
- **Debris**: in the aftermath of earthquakes or landslides, the presence of debris serves as evidence of collapsed buildings.
- **Flood**: in both natural and man-made disasters, detecting floods is crucial for assessing the severity of the situation and implementing effective mitigation measures.
- **Smoke**: while closely related to fire, smoke detection is also useful for events like explosions, fires, gas leakages, volcanic eruptions, or transportation accidents.
- **Person**: the presence of people in the disaster area, if visible from aerial views, is indicative of tentative victims (civil persons) or first-responders actively engaged in SAR missions. Identifying people in aerial images presents challenges, as they often appear in a few pixels, being difficult to detect and impossible to distinguish between a civil or a first-responder person.
- **Car**: crushed cars are potential locations of victims and are commonly found not only in traffic accidents but also in floods, landslides, and collapsed buildings, among others.

TABLE I: Results for aerial views with $IoU = 0{,}5$

Class	Labels	P (%)	R (%)	mAP@.5 (%)
pedestrian	7985	54.7	38.5	41.4
people	4735	49.0	32.8	33.5
bicycle	1064	29.6	14.8	14.1
car	12500	70.3	75.2	76.5
van	1819	41.3	37.8	37.0
truck	718	45.1	35.0	35.2
tricycle	964	46.1	14.0	17.1
awning-tricycle	486	27.5	12.1	7.97
bus	234	51.1	41.6	42.9
motor	4392	48.8	43.3	40.6
all	34897	46.3	34.5	34.6

- **Emergency Vehicle**: the detection of vehicles used for SAR tasks, if possible, offers valuable insights into the areas where rescuers are actively working. Although distinguishing between civilian cars and emergency vehicles in aerial views can be inappreciable, detecting trucks or another type of machinery is feasible.

The first four classes (fire, debris, flood, and smoke) fall into the named category of complex-shape classes, as they do not correspond to easily discernible objects in aerial views. These classes lack clear edges to define the shape of the target object and/or the object is not salient enough from the background. In Section IV we deeply analyze these challenging complex-shape classes.

IV. Experimental results

IV-A. Adapting the model for aerial images

As described in Section II, the model is firstly adapted to classify objects in aerial imagery using the VisDrone dataset. In this first step, we employed the existing VisDrone labels, which comprise classes such as pedestrian, people, bicycle, car, van, truck, tricycle, awning-tricycle, bus, and motor [18]. In addition, we fine-tuned the hyperparameters of the model for aerial views, following the approach proposed in [21]. These changes allow to train the model for more epochs without overfitting, enhancing its performance.

These results can be seen in Table I for each class, with the number of labels in the second column, and the mean Average Precision (mAP) for a 50 % IoU threshold in the last column. As can be seen, the mAP over all classes is shown at the bottom of the table, showing a moderate value of 34.6 %. Although this result may appear deficient, it is important to note that these results are comparable to the outcomes of other adapted models for aerial imagery in the VisDrone competition [18]. The relatively lower mAP can be attributed to the dataset imbalance, as evident from the second column in the table. For instance, the class car has 12.500 labels and achieves the best mAP of 76.5 %. By contrast, the awning-tricycle class comprises only 486 labels, leading to the lowest mAP (7.97 %) as a consequence. The class imbalance directly impacts the model's performance, with the most represented class, the class car in this case, benefiting from more extensive training data, resulting in superior performance compared to less represented classes. Thus, these results are consistent with the expected behavior

TABLE II: Results for SAR-classes with $IoU = 0{,}5$

Class	Labels	P (%)	R (%)	mAP@.5 (%)
car	3272	75.5	31.1	36.0
person	17702	82.6	84.6	86.7
smoke	2703	73.4	60.8	60.6
debris	5677	85.7	79.2	82.9
flood	1200	100.0	54.4	58.9
emergency-vehicle	1008	90.8	69.2	73.4
fire	1964	60.9	35.0	42.6
all	33528	81.3	59.2	63.0

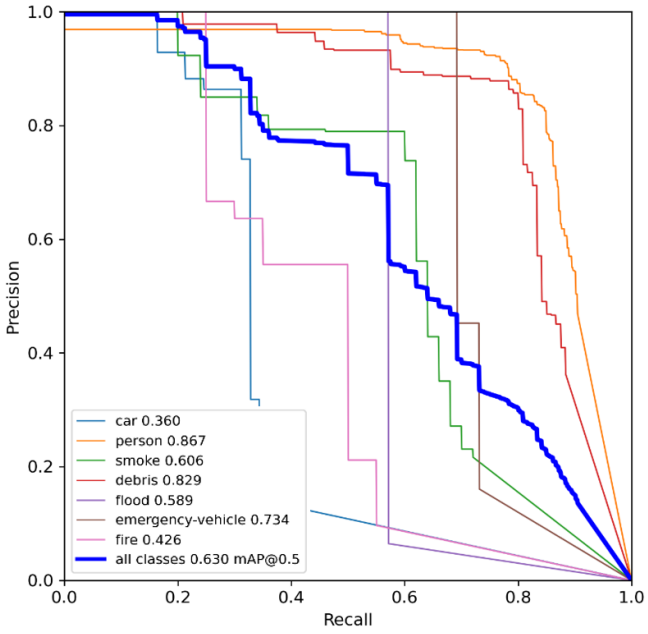

Fig. 4: Precision-recall curves for the SAR-related classes.

during the adaptation to aerial views using this dataset [21]. Class imbalance is a common challenge in object tasks, particularly in aerial scenarios, which is explored in the next stage of this study.

IV-B. Training in disaster scenarios

Once the weights of the model have been adapted for aerial imagery we proceed to optimize the model for SAR-related classes. As was explained in Section II, firstly we use the UMA-SAR dataset. To address the class imbalance in the training set, we augment the dataset by including carefully selected images from the AIDER dataset for oversampling the underrepresented classes and achieving a more balanced dataset. The AIDER images contain a higher prevalence of events like fire, smoke, and flood, making it a valuable resource for augmenting these specific classes in the training set. A total of 505 images from UMA-SAR and AIDER datasets were manually labeled. Finally, we employed another usual technique, data augmentation, to further enhance the model performances during optimization and also to address the class imbalance problem [22]. Specifically, we use rotation, flipping, scaling, changes in saturation and contrast on all images in the training set.

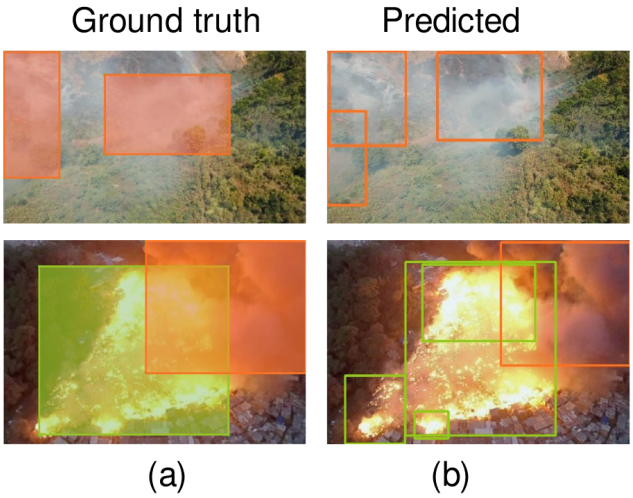

Fig. 5: Comparative between (a) labeled and (b) predicted bounding boxes for complex-shape classes. Fire (green) and smoke (orange) classes.

Table II presents the results of this optimization on SAR-related classes. It can be noted that in this optimization we achieve a better mAP of 63 %. Some classes like person, debris and, emergency-vehicle are above 70 % of mAP for the 50 % of the IoU threshold. In contrast, in this training set the car class is the least represented achieving just 36 % mAP as a result. The remaining SAR-related classes (smoke, flood, and fire) achieve average mAP values. Although the car class has more labels in the dataset than floods or smoke, the car class appears in only a few pixels in aerial images while floods or smoke often exhibit more extensive areas in an image.

Figure 4 shows the precision-recall (P-R) curves of the optimization results in the SAR-related classes. Although receiver operating characteristic (ROC) curves are insensitive to class imbalance, and the P-R curves are not, they are more informative in model performance than ROC [23]. The P-R curves for person and debris classes are close to an ideal P-R curve when both precision and recall reach 1.0. However, the car class has the poorest curve, with a precision dropping at low recall values. The remaining classes are close to the mean P-R curve for all classes, indicative of class imbalance [24]. In spite of this issue, high precision results imply a significant number of true detection [25]. We used data augmentation for all labeled images, but we may select more augmented operations in those images showing rare objects in order to obtain a more balanced training set.

In addition, complex-shape classes, such as debris, flood, fire, and smoke, pose challenges in the model performance due to differences between the labeled and predicted bounding boxes. The lack of clear edges in aerial imagery and/or the low background saliency make it difficult to define precise bounding boxes for these classes during manual labeling. Fig. 5 provides two illustrative examples of this ambiguity, where the labeled images for fire and smoke

TABLE III: Results for SAR-classes with IoU = 0.5 comparing optimizers

Optimizer	SGD		Adam	
Epochs	130	150	130	150
Class	mAP@.5 (%)		mAP@.5 (%)	
car	41.2	36.0	26.4	27.0
person	85.5	86.7	73.5	70.9
smoke	69.7	60.6	53.2	48.1
debris	80.2	82.9	75.0	76.0
flood	20.4	58.9	37.4	53.4
emergency-vehicle	75.0	73.4	55.6	52.3
fire	48.2	42.6	50.8	53.8
all	60.0	63.0	53.1	54.5

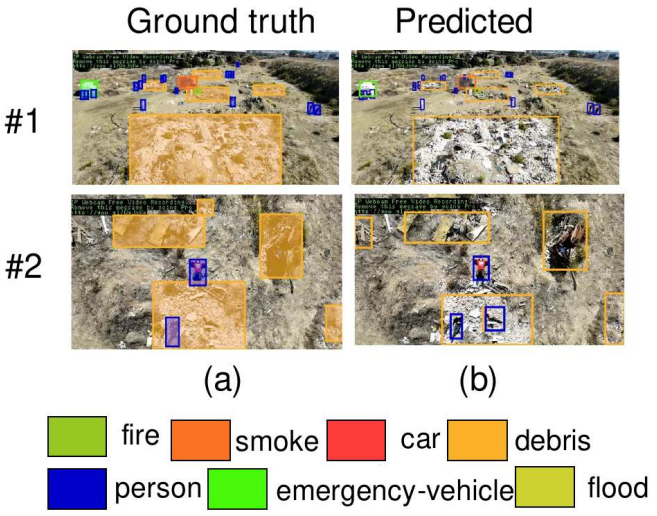

Fig. 6: Model detection in UMA-SAR dataset images.

classes (on the left) clearly differ from the predicted ones (on the right). As a result, the IoU values are low, leading to lower mAP for these classes. It is difficult to decide whether the labeled bounding boxes are more appropriate than the predicted ones.

An alternative study to test the model performance is to explore different optimizers and the number of training epochs. Table III presents the results obtained using both Stochastic Gradient Descent (SGD) and Adam optimizers, along with different training epochs. The best choice for our model is to use the SGD optimizer and 150 epochs.

Figure 6 displays some examples in the UMA-SAR dataset, demonstrating that the model is able to detect all classes, except for floods, due to the absence of instances in this dataset. The model effectively recognizes debris areas and even identifies victims among them, as can be seen in case #2, where a dummy is detected as a person, but it was not labeled. In these images, it can be noted how the detection of people is a challenge, due to their appearance in only a few pixels. On the other hand, Figure 7 illustrates images from the AIDER dataset, demonstrating the model's ability to detect catastrophic areas and its capacity to generalize across various situations. Despite the challenge of detecting floods due to corresponding to areas with very few features,

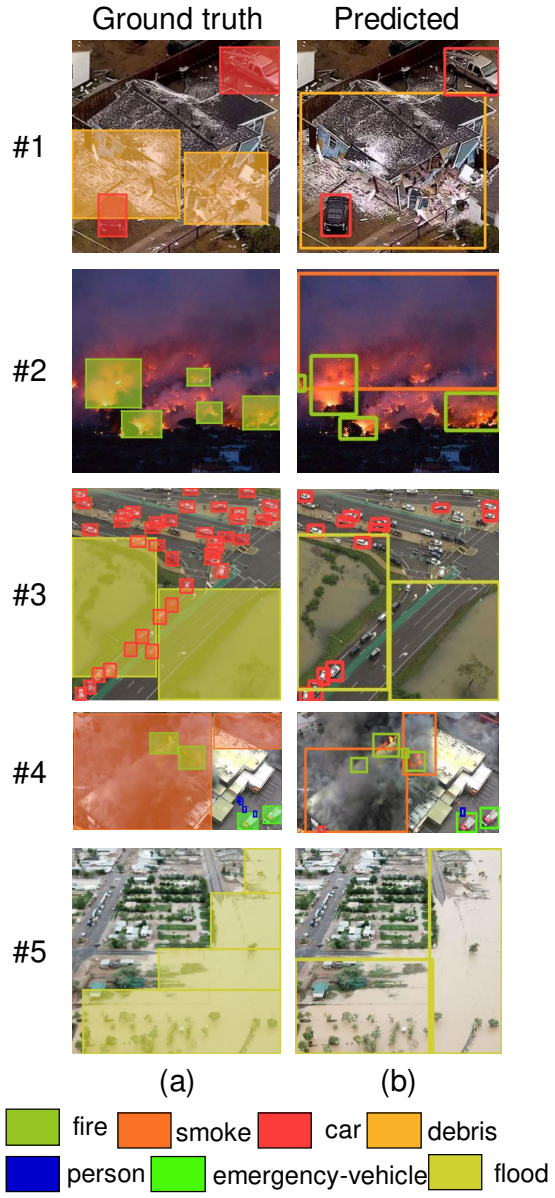

Fig. 7: Model detection in AIDER dataset images.

the model still demonstrates a remarkable ability to identify such areas, in cases #3 and #5. In cases #1 and #3 cars are correctly detected but in #3 only half of them are. Case #4 is illustrative of how emergency-vehicles are distinguished from civilian cars.

V. Conclusions

This paper presents a convolutional neural network model for disaster area recognition in aerial imagery. By adapting the YOLOv5 network for aerial images and optimizing it for detecting Search and Rescue (SAR) related classes, we have developed a fine-tuned model that effectively detects objects and events in disaster scenarios. In order to optimize the model for disaster areas, we created a training set labeling two publicly available SAR datasets, defining seven SAR-

related classes, such as person, car, debris, fire, smoke, flood, and emergency-vehicle. To address class imbalance, we labeled images from the UMA-SAR dataset complemented with some selected images from the AIDER dataset to oversample the underrepresented classes, in particular fire, smoke, and flood. Model optimization in SAR scenes achieves efficient performance with an optimal configuration using the SGD optimizer and 150 epochs for training. Finally, an analysis of the obtained results discusses the intricacies related to the so-called complex-shape classes, those with missing edges and/or background saliency. Although the model performance demonstrates promising outcomes, a more balanced SAR dataset is needed to improve detection in some SAR-related classes. In addition, exploring the use of semantic segmentation for SAR datasets in aerial imagery is another interesting future work for object detection at pixel level. However, semantic segmentation in aerial views poses problems because some classes only appear in a few pixels.

ACKNOWLEDGEMENTS

The authors want to thank the collaboration of the Chair for Safety, Emergencies and Disasters of the Universidad de Málaga, led by Jesús Miranda for organizing the exercises. In addition, we would like to thank all the members of the Robotics and Mechatronics Lab, for obtaining aerial images during SAR exercises. This work has been partially funded by the Ministerio de Ciencia, Innovación y Universidades, Gobierno de España, project PID2021-122944OB-I00.

REFERENCES

[1] B. Doroodgar, Y. Liu, and G. Nejat, "A learning-based semi-autonomous controller for robotic exploration of unknown disaster scenes while searching for victims," *IEEE Transactions on Cybernetics*, vol. 44, no. 12, pp. 2719–2732, 2014.

[2] F. Pastor, F. J. Ruiz-Ruiz, J. M. Gómez-de Gabriel, and A. J. García-Cerezo, "Autonomous wristband placement in a moving hand for victims in search and rescue scenarios with a mobile manipulator," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 11 871–11 878, 2022.

[3] J. Sánchez-García, J. M. García-Campos, M. Arzamendia, D. G. Reina, S. Toral, and D. Gregor, "A survey on unmanned aerial and aquatic vehicle multi-hop networks: Wireless communications, evaluation tools and applications," *Computer Communications*, vol. 119, pp. 43–65, 2018.

[4] M. Toscano-Moreno, J. Bravo-Arrabal, M. Sánchez-Montero, J. Serón-Barba, R. V. Martín, J. J. F. Lozano, A. Mandow, and A. García-Cerezo, "Integrating ROS and android for rescuers in a cloud robotics architecture: Application to a casualty evacuation exercise," in *IEEE International Symposium on Safety, Security, and Rescue Robotics, SSRR 2022, Sevilla, Spain, November 8-10, 2022*. IEEE, 2022, pp. 270–276. [Online]. Available: https://doi.org/10.1109/SSRR56537.2022.10018629

[5] R. Edlinger, C. Föls, and A. Nüchter, "An innovative pick-up and transport robot system for casualty evacuation," in *IEEE International Symposium on Safety, Security, and Rescue Robotics, SSRR 2022, Sevilla, Spain, November 8-10, 2022*. IEEE, 2022, pp. 67–73. [Online]. Available: https://doi.org/10.1109/SSRR56537.2022.10018818

[6] C. Kyrkou, "Aider: Aerial image database for emergency response applications," 2020. [Online]. Available: https://github.com/ckyrkou/AIDER

[7] C. Kyrkou and T. Theocharides, "Emergencynet: Efficient aerial image classification for drone-based emergency monitoring using atrous convolutional feature fusion," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 1687–1699, 2020.

[8] D. Broyles, C. R. Hayner, and K. 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)*, 2022, pp. 9467–9474.

[9] T. Chowdhury, R. Murphy, and M. Rahnemoonfar, "Rescuenet: A high resolution uav semantic segmentation benchmark dataset for natural disaster damage assessment," 2022.

[10] H. C. Reis and V. Turk, "Detection of forest fire using deep convolutional neural networks with transfer learning approach," *Applied Soft Computing*, vol. 143, p. 110362, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1568494623003800

[11] C. Bahhar, A. Ksibi, M. Ayadi, M. M. Jamjoom, Z. Ullah, B. O. Soufiene, and H. Sakli, "Wildfire and smoke detection using staged yolo model and ensemble cnn," *Electronics*, vol. 12, no. 1, 2023. [Online]. Available: https://www.mdpi.com/2079-9292/12/1/228

[12] X. Zhu, J. Liang, and A. Hauptmann, "Msnet: A multilevel instance segmentation network for natural disaster damage assessment in aerial videos," *Proceedings - 2021 IEEE Winter Conference on Applications of Computer Vision, WACV 2021*, pp. 2022–2031, 1 2021.

[13] Z. Hong, H. Zhong, H. Pan, J. Liu, R. Zhou, Y. Zhang, Y. Han, J. Wang, S. Yang, and C. Zhong, "Classification of building damage using a novel convolutional neural network based on post-disaster aerial images," *Sensors*, vol. 22, no. 15, 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/15/5920

[14] A. K. Jena, S. S. Potru, D. R. Balaji, A. Madu, and K. Chaurasia, "Disaster risk mapping from aerial imagery using deep learning techniques," in *Proceedings of UASG 2021: Wings 4 Sustainability*, K. Jain, V. Mishra, and B. Pradhan, Eds. Cham: Springer International Publishing, 2023, pp. 319–329.

[15] S. M. Ali Rizvi, R. Mobin Ahmed, K. G. Alamdar, M. Raza Khorasani, and J. A. Memon, "Human detection and localization in indoor disaster environments using uavs," in *2022 4th International Conference on Robotics and Computer Vision (ICRCV)*, 2022, pp. 159–163.

[16] P. Byukusenge and Y. Zhang, "Life detection based on uavs - thermal images in search and rescue operation," in *2022 IEEE 22nd International Conference on Communication Technology (ICCT)*, 2022, pp. 1728–1731.

[17] S. Bae, H. Shin, H. Kim, M. Park, M.-Y. Choi, and H. Oh, "Deep learning-based human detection using rgb and ir images from drones," *International Journal of Aeronautical and Space Sciences*, pp. 2093–2480, 2023.

[18] P. 2022, "Aerial-ground intelligent unmanned system environment perception challenge." [Online]. Available: http://aiskyeye.com/

[19] J. Morales, R. Vázquez-Martín, A. Mandow, D. Morilla-Cabello, and A. García-Cerezo, "The UMA-SAR dataset: Multimodal data collection from a ground vehicle during outdoor disaster response training exercises," *The International Journal of Robotics Research*, vol. 40, no. 6-7, pp. 835–847, 2021.

[20] J. Morales, R. Vázquez-Martín, A. Mandow, D. Morilla-Cabello, and A. García-Cerezo, "The UMA-SAR Dataset: Multimodal data collection from a ground vehicle during outdoor disaster response training exercises," *The International Journal of Robotics Research*, vol. 40, no. 6-7, pp. 835–847, 2021.

[21] X. Zhu, S. Lyu, X. Wang, and Q. Zhao, "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*, 2021, pp. 2778–2788.

[22] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[23] M. R. Takaya Saito, "The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets," *PLoS ONE*, vol. 10, no. 3, pp. 1–21, 2020.

[24] C. K. Williams, "The effect of class imbalance on precision-recall curves," *Neural Computation*, vol. 33, no. 4, pp. 853–857, 2020.

[25] K. P. Murphy, *Probabilistic Machine Learning: An introduction*. MIT Press, 2022. [Online]. Available: probml.ai