Advancing Off-Road Operations: Comparative Assessment of Object Detection Methodologies

Javier Saez-Perez

University of the West of Scotland
Paisley, United Kingdom
0000-0003-2114-9798
javier.saez-perez@uws.ac.uk

Qi Wang

University of the West of Scotland
Paisley, United Kingdom
0000-0002-7764-9858
qi.wang@uws.ac.uk

Jose Maria Alcaraz Calero

University of the West of Scotland
Paisley, United Kingdom
0000-0002-2654-7595
jose.alcaraz-calero@uws.ac.uk

Jose Garcia-Rodriguez

University of Alicante
Alicante, Spain
0000-0002-7798-3055
jgarcia@dtic.ua.es

***Abstract*—Off-road operations such as search and rescue or goods delivery in remote wilderness environments demand efficient and effective strategies to locate individuals. Traditional Search and Rescue (SAR) methods, such as canine units and human-assisted search, face challenges in timeliness, cost, scalability and adaptability. To address these limitations, we explore the integration of ground robots equipped with LiDARs into SAR workflows, focusing on off-road scenarios where adverse weather or complex terrain hinder conventional methods like Unmanned Aerial Vehicles (UAVs). Our study conducts a comprehensive analysis of single-sensor LiDAR algorithms for small object detection in off-road environments, utilizing a newly curated dataset tailored for SAR applications. We identify the critical role of sensor configuration and environmental factors in detection performance, providing insights for designing robust detection systems. By filling gaps in existing literature and elucidating the strengths of single-sensor LiDAR algorithms, we contribute to advancing off-road operations methodologies and technology. Our findings underscore the potential of integrating ground robots into off-road operations and offer valuable guidance for enhancing search capabilities in challenging environments.**

***Index Terms*—object detection, artificial intelligence, 3D perception, robotics**

I. Introduction

Off-road operations teams are in a perpetual quest to refine their operational strategies. For example, for Search and Rescue (SAR) operations, the teams are constantly exploring innovative techniques to expedite the location of individuals lost in remote wilderness environments. Traditionally, canine units have been the frontline resource in these endeavors. However, the efficacy of this method is tempered by the time investment required for training and deployment, as well as the logistical constraints associated with coordinating multiple dog teams. As such, a pressing need exists for alternative approaches that can supplement or even supplant traditional search methods. Human-assisted search methodologies represent one such avenue, relying on the concerted efforts of SAR personnel, volunteers, and community members to scour expansive areas in search of the lost individual. While this approach holds promise in its potential to leverage human intuition and adaptability, it is inherently labor-intensive and dependent on the availability of sufficient manpower [1]. Consequently, there is an ongoing drive within the SAR community to explore technological solutions that can enhance search efficiency and effectiveness, particularly in scenarios where time is of the essence and lives hang in the balance. In recent years, the advent of Unmanned Aerial Vehicles (UAVs) has revolutionized SAR operations by providing aerial reconnaissance capabilities. UAVs equipped with high-resolution cameras transmit real-time images of the search area, enabling SAR teams to cover vast expanses of terrain efficiently. Furthermore, the application of object detection algorithms on the images captured by UAVs allows for automated analysis and identification of potential targets, further streamlining the search process and increasing the likelihood of locating missing individuals. This integration of UAV technology with advanced object detection algorithms represents a significant advancement in SAR operations, offering a cost-effective and scalable solution to the challenges posed by remote wilderness environments.

Nonetheless, alternative strategies must be considered to augment SAR operations in challenging scenarios where weather conditions, such as poor visibility or strong winds, impede the use of UAVs. Ground robots emerge as a viable solution in such circumstances, offering the capability to access remote and difficult terrains where human intervention may be impractical or unsafe. Leveraging ground robots equipped with multiple sensors, SAR teams can navigate rugged landscapes and search areas inaccessible to humans, thereby expanding the scope and effectiveness of search efforts. Despite their potential, there remains a notable scarcity of research exploring the integration of ground robots into SAR operations, particularly in the context of off-road scenarios. Additionally, existing literature on SAR often overlooks the utilization of multi-sensor systems, which have demonstrated efficacy in diverse environments but remain under-explored in off-road settings. Addressing these gaps in research presents an opportunity to enhance SAR capabilities by leveraging the complementary strengths of ground robots and multi-sensor systems to improve detection and localization of missing persons in challenging conditions.

In summary, the main contributions of this study are as follows:

- We explore the underutilized potential of ground robots in SAR operations. By assessing the viability of integrating ground robots equipped with single-sensor LiDAR into SAR workflows, we aim to evaluate its search capabilities and effectiveness in challenging environments.
- We conduct a comprehensive analysis of single-sensor LiDAR algorithms for small object detection in off-road scenarios. Leveraging a newly curated dataset tailored for SAR applications, we systematically evaluate the efficacy of each algorithm across diverse environmental contexts. This thorough evaluation sheds light on the critical role of sensor configuration and environmental factors in detection performance, providing valuable insights for designing robust detection systems for real-world SAR applications.
- Our research fills notable gaps in existing literature by focusing on off-road scenarios and the integration of ground robots equipped with single-sensor LiDAR into SAR operations. By elucidating the strengths and limitations of single-sensor LiDAR algorithms in improving detection and localization of missing persons, especially in challenging conditions, we contribute to the knowledge of room for improvement and advancement of SAR methodologies and technology.

This paper is structured as follows: Section 2 reviews related work in small 3D object detection. Section 3 describes the datasets used in the experiments. Section 4 presents 3D object detection algorithms used in the experiments. Section 5 outlines our experimental results and analysis. Last, in Section 6, the conclusions of the investigation are presented and discussed.

II. Related Work

Off-road operations such as SAR operations in remote wilderness environments have historically relied on traditional methods, including canine units and human-assisted search methodologies. However, with the emergence of technological innovations, there has been a notable shift towards integrating advanced solutions to enhance search efficiency and effectiveness. In recent years, Unmanned Aerial Vehicles (UAVs) have revolutionized SAR operations by providing aerial reconnaissance capabilities, allowing SAR teams to cover vast areas efficiently and employ object detection algorithms for automated analysis. A comparative analysis of published results reveals a spectrum of approaches and methodologies aimed at optimizing UAV-based SAR missions. For instance, [1] employed Googlenet and Support Vector Machine (SVM) algorithms to locate avalanche victims, albeit with computational constraints unsuitable for constrained environments. Similarly, [2] proposed an SVM-based human detector, emphasizing Histograms of Oriented Gradients (HOG) features, yet facing challenges in computational intensity. [3] reported a sensitivity of 76.97% in UAV-based SAR missions, albeit with altitude limitations. [4] utilized Mobilenet for human detection and respiration monitoring, though lacking specificity on detection speed and limited applicability for small object detection. [5] demonstrated the utilization of POINet (Mobilenetv2) but faced constraints in processing speed and altitude coverage. [6] achieved high accuracy in detecting individuals in aerial imagery, yet real-time performance remains a focus for future research.

While UAVs have demonstrated significant promise, challenges persist in scenarios where adverse weather conditions hinder their use. To address these challenges, researchers have explored alternative solutions, including the integration of ground robots equipped with multiple sensors and collaboration with UAVs. Ground robots offer the capability to access rugged terrains inaccessible to humans, expanding the scope of search efforts. However, there remains a dearth of research exploring the integration of ground robots. They are usually equipped with one or more sensors. These sensors go from cameras to LiDAR. In this work, we focus on 3D object detection in point clouds.

The field of 3D object detection has witnessed remarkable progress in recent years, driven by innovative methodologies and novel architectures proposed in various research endeavors. Among these, VoteNet [7] stands out for its groundbreaking approach, as delineated in its corresponding paper. By transmuting depth images into point clouds and embedding Hough voting within an end-to-end differentiable architecture, VoteNet has emerged as a frontrunner in 3D object detection, showcasing unparalleled performance on benchmark datasets like SUN RGB-D and ScanNetV2. While its achievements are noteworthy, the sparse nature of data presents a persistent challenge, particularly in accurately regressing 3D object centroids, and its confinement solely to point cloud data, excluding image inputs, delineates its limitations. Similarly, Shape Signature Networks (SSN) [8] heralds a new era in 3D object detection by introducing a unique shape signature for enhanced multi-class discrimination in point cloud analysis. Leveraging symmetry, convex hull, and Chebyshev fitting operations, SSN has attained state-of-the-art results on extensive datasets such as nuScenes [9] and Lyft [10], underscoring its efficacy in addressing the intricacies of 3D shape information representation. Its contribution lies not only in the compact and robust encoding of shape features but also in the adept handling of class-specific distinctions, thus mitigating the challenges associated with accurate 3D object detection. In a divergent trajectory, the Single-stage Monocular 3D Object Detection from KeyPoints (SMOKE) [11] methodology presents a paradigm shift by circumventing traditional 2D detection stages. Through a multi-step disentanglement approach, SMOKE has demonstrated remarkable improvements in both accuracy and speed, eclipsing existing monocular 3D detection methods on prominent datasets such as KITTI [12]. Its innovative approach to 3D bounding box regression and strategic emphasis on training convergence augmentation underscore its significance in pushing the boundaries of monocular 3D object detection. Concurrently, the RegNet paper [13] delves deep into the intricacies of network design, offering insights into the RegNet models that outperform their counterparts like EfficientNet in terms of speed and performance. By parameterizing networks and meticulously analyzing the design space, the paper presents a promising trajectory for future research, wherein network structure, width, depth, and grouping dynamics are meticulously orchestrated for optimal performance. Further enriching the landscape of 3D object detection are methodologies such as PointPillars [14], Position Adaptive Convolution (PAConv) [15], and CenterPoint [16], each offering distinctive contributions to the field. PointPillars, with its novel trade-off between speed and accuracy, embodies the essence of end-to-end learning through 2D convolutional layers. PAConv, on the other hand, introduces a versatile convolutional operation tailored explicitly for point cloud tasks, thereby enhancing performance across various domains. Meanwhile, CenterPoint pioneers a center-based framework, simplifying 3D object detection and tracking while delivering state-of-the-art results on challenging benchmarks like Waymo and nuScenes [9].

In summation, these seminal contributions underscore the vibrant ecosystem of 3D object detection research, characterized by incessant innovation and iterative refinement. As researchers continue to push the boundaries of what's possible, leveraging diverse methodologies and architectures, the future of 3D object detection appears promising, with each stride forward bringing us closer to realizing its full potential in real-world applications.

III. Dataset

In this study, we leveraged our off-road dataset to comprehensively investigate our research objectives, offering unique insights into different environmental contexts. This dataset aligns with the use case of off-road applications.

A pioneering resource tailored for off-road operations, meticulously compiled to drive advancements in search and rescue missions, goods delivery, and beyond. This groundbreaking dataset captures the essence of diverse off-road scenarios, blending the rugged landscapes of Alicante, Spain, with the dynamic terrain of Paisley, United Kingdom, offering researchers a multifaceted lens into off-road challenges.

At the heart of this dataset lies a sophisticated sensor ensemble, comprising the Velodyne VLP-32C LiDAR and the ZED 2 stereo camera, meticulously mounted on a robust frame atop the vehicle. This setup ensures comprehensive coverage and depth perception, empowering researchers to navigate and understand the intricacies of off-road environments with unparalleled precision.

Comprising 2500 training examples and 600 testing examples, this dataset provides a rich tapestry of off-road scenarios, each meticulously annotated to focus on a singular class: humans. By honing in on this pivotal class, researchers can delve deep into human detection algorithms, paving the way for enhanced safety and efficiency in off-road operations.

The diversity of locations—from the picturesque landscapes of Alicante to the rugged terrains of Paisley—imbues the dataset with a wealth of variability, enriching research endeavors and ensuring robustness in algorithmic development. From navigating through dense forests to traversing rocky terrains, the Off-Road Dataset offers a comprehensive spectrum of challenges, pushing the boundaries of off-road robotics and autonomy.

As a testament to its utility and versatility, the Off-Road Dataset serves as a catalyst for innovation across a myriad of applications, including autonomous exploration, disaster response, and remote logistics. By fostering collaboration and pushing the boundaries of off-road capabilities, this dataset empowers researchers to unlock new horizons in the realm of intelligent systems.

Researchers eager to harness the potential of the Off-Road Dataset can embark on their journey of exploration and discovery by accessing the dataset through the provided channels. Through relentless innovation and collaboration, the Off-Road Dataset charts a course towards a future where off-road operations are safer, more efficient, and more accessible than ever before.

IV. Models

This section presents the models' architectures that were used during the experiments.

A. Part-A2

The Part-A2net model (see figure 1) represents a significant advancement in the field of 3D object detection from LiDAR point cloud data. This innovative framework is structured around two key stages: the part-aware stage-I and the part-aggregation stage-II, each playing a crucial role in enhancing the accuracy and efficiency of object detection tasks. In the part-aware stage-I, the model leverages free-of-charge intra-object location labels and foreground labels derived from ground-truth 3D box annotations to predict precise intra-object part locations. By incorporating this information, the model not only generates high-quality 3D proposals but also improves the overall quality of predicted bounding boxes.

A notable feature of the Part-A2net model is the introduction of two distinct strategies for 3D proposal generation: the anchor-free scheme and the anchor-based scheme. The anchor-free approach offers a lightweight and memory-efficient solution by directly generating 3D bounding box proposals through foreground point segmentation and proposal generation simultaneously. In contrast, the anchor-based strategy utilizes pre-defined 3D anchor boxes on downsampled bird-view feature maps to achieve higher recall rates, albeit with increased memory and computational requirements. This flexibility in proposal generation strategies allows the model to adapt to different detection scenarios and optimize performance based on specific requirements.

The RoI-aware point cloud pooling module, a novel component of the Part-A2net framework, plays a pivotal role in grouping predicted intra-object part locations and encoding geometry-specific features of each 3D proposal. By normalizing diverse 3D proposals to a unified coordinate system, this module effectively captures geometric information and enhances feature encoding, thereby improving the model's ability to discriminate between object classes and accurately localize objects in complex scenes.

Moreover, the part-aggregation stage-II refines box locations by exploring the spatial relationships of pooled intra-object part locations. This stage focuses on capturing the intricate geometric details of object parts to accurately score the boxes and refine their positions, leading to enhanced detection accuracy and robustness. The comprehensive design of the Part-A2net

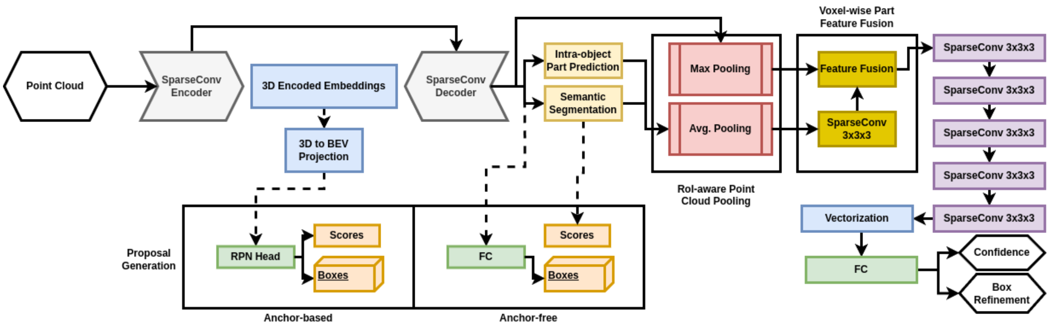

Fig. 1. Architecture of the Part-A2 model presented in [17].

model, coupled with its innovative components and advanced functionalities.

B. *Point Pillars*

The PointPillars model (see figure 2) presents a pioneering approach to encoding point clouds for downstream object detection tasks. Departing from traditional fixed encoders or data-learned encoders, PointPillars harnesses PointNets to derive a representation of point clouds structured in vertical columns, referred to as pillars. This novel architecture enables the model to fully capture the information embedded in the point cloud data, empowering it to directly learn features from the data without dependence on pre-defined encodings. Operating on pillars rather than voxels, PointPillars obviates the need for manual tuning of vertical direction binning, rendering the process notably more efficient and effective.

One of the paramount advantages of PointPillars lies in its adeptness at striking a delicate balance between speed and accuracy, substantially outperforming previous encoders in both aspects. By extracting features directly from the point cloud data, PointPillars can exploit the entirety of the information encapsulated within the point cloud, thereby enhancing accuracy in object detection tasks. Moreover, the model's efficient computation on GPUs, facilitated by formulating critical operations as 2D convolutions, further bolsters its performance and speed. PointPillars operates at a remarkable speed of 62 Hz.

V. Experiments and Results

A. *Experiments*

The algorithms were trained and evaluated on a machine equipped with an Intel i7 CPU, 32 GB of RAM memory, and an NVIDIA GTX 1080 GPU on Ubuntu 20.04. This hardware configuration provided sufficient computational resources for training and inference tasks. We utilized the Off-Road Dataset described in Section III for training and evaluation purposes. This dataset, specifically tailored for off-road applications, encompasses diverse environmental contexts, providing a comprehensive testbed for assessing algorithm performance. For evaluation, we adopted Mean Average Precision (mAP) as our primary metric, a standard measure in object detection tasks. We focused on the detection of people within the off-road environment, aiming to assess the algorithms' ability to accurately localize and classify human subjects.

All the models were trained using a cyclic learning rate schedule along with the AdamW optimizer, starting at 0.0018 and increasing to 10 times that value over 16 epochs, then exponentially decreasing to 0.000018 over the next 24 epochs. We employed the AdamW optimizer with momentum adjustments following a similar cosine annealing schedule.

Our evaluation focused on two distinct modes: 3D bounding box (bbox) detection and Bird's Eye View (BEV).

- **3D Bounding Box Detection:** In this mode, the algorithms are tasked with detecting objects in the full 3D space, including their precise position, size, and orientation relative to the sensor. This mode is crucial for SAR operations, where accurately locating individuals within complex terrain is essential.
- **Bird's Eye View (BEV):** Unlike 3D bbox detection, BEV provides a top-down perspective of the environment, akin to viewing the scene from directly above. In BEV mode, objects are represented as 2D projections on a horizontal plane, simplifying the detection task by eliminating depth information. This mode is particularly useful for navigation and obstacle avoidance in off-road environments.

In the dataset evaluation, objects are classified into three difficulty levels: Easy, Moderate, and Hard. Easy objects have a minimum bounding box height of 40 pixels, maximum occlusion level of fully visible, and maximum truncation of 15%. Moderate objects have a minimum bounding box height of 25 pixels, maximum occlusion level of partly occluded, and maximum truncation of 30%. Hard objects have a minimum bounding box height of 25 pixels, maximum occlusion level of difficult to see, and maximum truncation of 50%. These difficulty levels serve as benchmarks for evaluating algorithm performance, with all methods ranked based on moderately

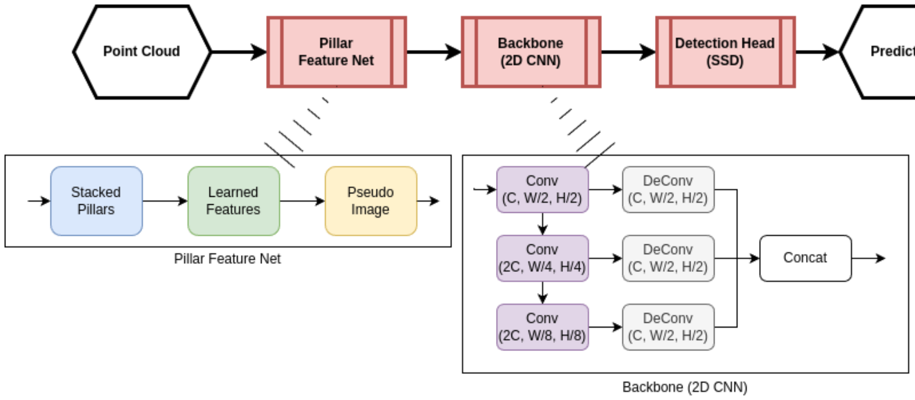

Fig. 2. Architecture of the Point Pillars model presented in [14].

TABLE I
PERFORMANCE COMPARISON OF OBJECT DETECTION ALGORITHMS IN OFF-ROAD ENVIRONMENTS: MEAN AVERAGE PRECISION (MAP) AT DIFFERENT INTERSECTION OVER UNION (IOU) THRESHOLDS AND EVALUATION MODES

			Point Pillars			Part-A2		
Metric	**IoU**	**Mode**	**Easy**	**Moderate**	**Hard**	**Easy**	**Moderate**	**Hard**
mAP	**0.5**	**3D**	47.30	47.15	47.15	34.81	32.96	32.96
		BEV	60.37	60.14	60.14	55.70	53.73	53.73
	0.25	**3D**	62.14	61.67	61.67	62.94	62.90	62.90
		BEV	63.51	62.32	62.32	63.37	63.18	63.18

difficult results.

B. Results

Our evaluation primarily focuses on mAP at an Intersection over Union (IoU) threshold of 0.5, a standard metric in object detection tasks. When examining the performance of Point Pillars, we observe mAP values ranging from 47.15% to 47.30% in the 3D mode and from 60.14% to 60.37% in the Bird's Eye View (BEV) mode across varying difficulty levels. Conversely, Part-A2 exhibits mAP values ranging from 32.96% to 34.81% in the 3D mode and from 53.73% to 55.70% in the BEV mode. Notably, Part-A2 consistently outperforms Point Pillars across both evaluation modes and difficulty levels, showcasing its efficacy in achieving higher detection accuracy.

We further investigate the performance of Point Pillars and Part-A2 at a lower IoU threshold of 0.25 to assess their robustness under stricter evaluation conditions. In the 3D mode, Point Pillars achieves mAP values ranging from 61.67% to 62.14%, while Part-A2 demonstrates comparable performance with mAP values ranging from 61.67% to 62.14%. Similarly, in the BEV mode, Point Pillars yields mAP values ranging from 62.32% to 63.51%, whereas Part-A2 achieves mAP values ranging from 62.90% to 63.37%. These results reaffirm the superior performance of Part-A2 relative to Point Pillars, underscoring its consistency across different evaluation scenarios.

The implications of our findings extend beyond mere performance metrics, offering valuable insights into the underlying mechanisms driving the efficacy of object detection methods. The observed superiority of Part-A2 can be attributed to several factors, including architectural design choices, optimization techniques, and data representation strategies. Moreover, our study sheds light on the nuanced interplay between evaluation metrics, IoU thresholds, and difficulty levels, highlighting the need for comprehensive benchmarking frameworks in evaluating object detection algorithms.

In conclusion, our comparative analysis provides a comprehensive understanding of the strengths and limitations of contemporary object detection methodologies. The findings presented herein serve as a stepping stone for further research endeavors aimed at advancing the state-of-the-art in object detection, with practical implications for real-world applications in autonomous systems, robotics, and beyond.

VI. CONCLUSIONS

In this study, we conducted a comprehensive investigation into the performance of two state-of-the-art object detection methodologies, namely Point Pillars and Part-A2, within the context of off-road environments. Leveraging our Off-Road Dataset, meticulously curated to emulate diverse off-road scenarios, we evaluated the efficacy of these methods across various evaluation metrics and scenarios.

Our experimental results reveal significant disparities in performance between Point Pillars and Part-A2 across different evaluation conditions. When assessing mAP at an IoU threshold of 0.5, Part-A2 consistently outperformed Point Pillars across both evaluation modes (3D and Bird's Eye View) and difficulty levels. This superiority was further underscored when evaluating at a lower IoU threshold of 0.25, reaffirming the robustness and efficacy of Part-A2 in achieving higher detection accuracy in off-road environments.

The observed performance differentials can be attributed to several factors, including architectural design choices, optimization techniques, and data representation strategies inherent to each methodology. Notably, the innovative features embedded within the Part-A2 framework, such as the RoI-aware point cloud pooling module and the refined part aggregation stage, contributed to its advantageous performance and robustness.

Our findings hold significant implications for the advancement of object detection methodologies in off-road applications. By shedding light on the strengths and limitations of contemporary approaches, our study provides valuable insights for researchers and practitioners in the field of autonomous systems, robotics, and beyond. In future work, continued research and development efforts are warranted to further enhance the performance and adaptability of object detection methodologies in challenging off-road environments.

In conclusion, our comparative analysis serves as a pivotal step towards the realization of safer, more efficient, and more accessible off-road operations, facilitated by the innovative capabilities of Part-A2 and its ilk. Through collaborative endeavors and continued innovation, we aim to propel the field of off-road robotics and autonomy towards new frontiers, guided by the principles of efficiency, reliability, and safety.

Acknowledgment

This work was in part funded by the EU Horizon Europe INCODE ("Programming platform for intelligent collaborative deployments over heterogeneous edge-IoT environments") project under the Grant number HORIZON-CL4-2022-DATA-01-03/101093069, EU Horizon 2020 ARCADIAN-IoT project ("Autonomous Trust, Security and Privacy Management Framework for IoT") under the Grant number H2020-SU-DS-2020/101020259 and the EU Horizon 2020 5G-PPP 5G-INDUCE project ("Open cooperative 5G experimentation platforms for the industrial sector NetApps") under the Grant number H2020-ICT-2020-2/101016941. The authors would like to thank all the partners in these projects for their support. The authors would like to acknowledge Thales UK for their support.

References

[1] M. B. Bejiga, A. Zeggada, and F. Melgani, "Convolutional neural networks for near real-time object detection from uav imagery in avalanche search and rescue operations," in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2016, pp. 693–696.

[2] A. Al-Kaff, F. M. Moreno, A. de la Escalera, and J. M. Armingol, "Intelligent vehicle for search, rescue and transportation purposes," in *2017 IEEE International Symposium on Safety, Security and Rescue Robotics (SSRR)*, 2017, pp. 110–115.

[3] W. G. Aguilar, M. A. Luna, J. F. Moya, V. Abad, H. Parra, and H. Ruiz, "Pedestrian detection for uavs using cascade classifiers with meanshift," in *2017 IEEE 11th International Conference on Semantic Computing (ICSC)*, 2017, pp. 509–514.

[4] A. Ivanovs, A. Nikitenko, M. Di Castro, T. Torims, A. Masi, and M. Ferre, "Multisensor low-cost system for real time human detection and remote respiration monitoring," in *2019 Third IEEE International Conference on Robotic Computing (IRC)*, 2019, pp. 254–257.

[5] R. Geraldes, A. Gonçalves, T. Lai, M. Villerabel, W. Deng, A. Salta, K. Nakayama, Y. Matsuo, and H. Prendinger, "Uav-based situational awareness system using deep learning," *IEEE Access*, vol. 7, pp. 122 583–122 594, 2019.

[6] D. Božić-Štulić, Ž. Marušić, and S. Gotovac, "Deep learning approach in aerial imagery for supporting land search and rescue missions," *International Journal of Computer Vision*, vol. 127, no. 9, pp. 1256–1278, Sep 2019. [Online]. Available: https://doi.org/10.1007/s11263-019-01177-1

[7] C. R. Qi, O. Litany, K. He, and L. J. Guibas, "Deep hough voting for 3d object detection in point clouds," 2019.

[8] X. Zhu, Y. Ma, T. Wang, Y. Xu, J. Shi, and D. Lin, "Ssn: Shape signature networks for multi-class object detection from point clouds," 2020.

[9] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuscenes: A multimodal dataset for autonomous driving," 2020.

[10] J. Houston, G. Zuidhof, L. Bergamini, Y. Ye, L. Chen, A. Jain, S. Omari, V. Iglovikov, and P. Ondruska, "One thousand and one hours: Self-driving motion prediction dataset," 2020.

[11] Z. Liu, Z. Wu, and R. Tóth, "Smoke: Single-stage monocular 3d object detection via keypoint estimation," 2020.

[12] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, Aug. 2013.

[13] I. Radosavovic, R. P. Kosaraju, R. Girshick, K. He, and P. Dollár, "Designing network design spaces," 2020.

[14] A. H. Lang, S. Vora, H. Caesar, L. Zhou, J. Yang, and O. Beijbom, "Pointpillars: Fast encoders for object detection from point clouds," 2019.

[15] M. Xu, R. Ding, H. Zhao, and X. Qi, "Paconv: Position adaptive convolution with dynamic kernel assembling on point clouds," 2021.

[16] T. Yin, X. Zhou, and P. Krähenbühl, "Center-based 3d object detection and tracking," 2021.

[17] S. Shi, Z. Wang, J. Shi, X. Wang, and H. Li, "From points to parts: 3d object detection from point cloud with part-aware and part-aggregation network," 2020.