Optimizing Object Detection for Maritime Search and Rescue: Progressive Fine-Tuning of YOLOv9 with Real and Synthetic Data

Luciano Lima[1], Fabio Andrade[1], Youcef Djenouri[1], Carlos Pfeiffer[2] and Marcos Moura[1]
[1]*University of South-Eastern Norway, Vestfold, Norway*
[2]*University of South-Eastern Norway, Porsgrunn, Norway*
{*limaluciano, fabio*}*@ieee.org,* {*youcef.djenouri, carlos.pfeiffer, marcos.moura*}*@usn.no*

Keywords: UAV, Transfer Learning, YOLOv9, Synthetic Data.

Abstract: The use of unmanned aerial vehicles for search and rescue (SAR) brings a series of advantages and reduces the time required to find survivors. It is possible to use computer vision algorithms to automate person detection, enabling a faster response from the rescue team. A major challenge in training image detection systems is the availability of data. In the SAR context, it can be more challenging as datasets are scarce. A possible solution is to use a virtual environment to generate synthetic data, which can provide an almost unlimited amount of data already labeled. In this work, the use of real and synthetic data for training the model YOLOv9t in maritime search and rescue operations is explored. Different proportions of real data were used for training a model from the scratch and for transfer learning by fine-tuning the model after being pretrained with synthetic data generated in Unreal Engine 4, to evaluate the performance aiming to reduce the reliance on real-world datasets. The total amounts of real and synthetic data were kept the same to ensure fair comparison. Fine-tuning a model pretrained on synthetic data with just 10% real data improved performance by 13.7% compared to using real data alone. An important finding is that the best performance was achieved with 70‘% real data instead a model trained solely on 100‘% real data. These results show that combining synthetic and real data enhances detection accuracy while reducing the need for large real-world datasets.

SCIENCE AND TECHNOLOGY PUBLICATIONS

1 INTRODUCTION

In Search and Rescue (SAR) missions performed in a maritime environment, the large area to be swept results in a high demand for personnel and a long duration to complete the mission. The rapid advance of technology enables artificial intelligence capabilities in small computers that can be embedded in unmanned aerial vehicles (UAV). While in regular search and rescue operations, the sweep is conducted by the crew of aircraft over the estimated location of distress (IAMSAR, 2022), the use of a swarm of drones can reduce both the time and cost of this operation. A key factor in a SAR mission is the time to reach to the survivors(Tuśnio and Wróblewski, 2022), what is related to reducing the time to detect the victim. A crucial stage of the UAV search effort is to recognize survivors among the debris in the sea. The ability to recognize images, as with any other machine learning technique, is highly dependent on a set of training images representing the target scenario (Géron, 2017). The efforts to acquire images of a search and rescue scenario at sea are challenging since it requires the mobilization of people, money, and vehicles. The use of images generated in a virtual environment with high visual fidelity can be a game-changer in training and improving the performance of survivor detection algorithms. In (Dabbiru et al., 2023), the Keras-RetinaNet framework was used to process synthetic aerial images generated in a simulated environment for building detection, showing promising results. In (Lima et al., 2023), the use of YOLOv8 in synthetic and real datasets was evaluated with the aim of validating the use of synthetic data to replace real-world data in training an object detection algorithm. However, the results showed that the presence of synthetic data reduced the system's accuracy, with no clear explanation for this. In the present work, the goal is to further develop this research using the best model containing synthetic data to find the optimal balance between synthetic and real images. The research developed in (Krump and Stütz, 2021) analyzes vehicle recognition using both real and synthetic images to determine which image descriptors have a higher influence on detection systems across both domains. Focusing in a better understand-

Lima, L., Andrade, F., Djenouri, Y., Pfeiffer, C. and Moura, M.
Optimizing Object Detection for Maritime Search and Rescue: Progressive Fine-Tuning of YOLOv9 with Real and Synthetic Data.
DOI: 10.5220/0013112700003890
Paper published under CC license (CC BY-NC-ND 4.0)
In *Proceedings of the 17th International Conference on Agents and Artificial Intelligence (ICAART 2025) - Volume 3*, pages 209-216
ISBN: 978-989-758-737-5; ISSN: 2184-433X
Proceedings Copyright © 2025 by SCITEPRESS – Science and Technology Publications, Lda.

ing of the reality gap, which is the difference between real and synthetic data characteristics, this study also identifies which image descriptors most significantly impact the generation of false positives when using YOLOv3. In (Jayalath and Munasinghe, 2021), a drone with an embedded image recognition system developed with Fast RCNN is used for SAR missions. A Raspberry Pi(Upton and Halfacree, 2014) was used for local computation on the drone. The scenario takes place over varied terrain, where an autonomous path planning system was developed to enable the drone to fly towards a suspicious target and send better images to an operator for confirming the presence of humans. The demand for customized datasets has already benefited from the advent of diffusion models. The use of text-to-image diffusion models to create image datasets with synthetic data was presented in (Xing and Tzes, 2023). YOLOv7 was employed to detect the presence of drones in various images.

2 BACKGROUND AND RELATED WORK

During a SAR operation at sea, the area to be swept is divided into a grid of cells, each with an assigned probability of containing survivors. This probability is assigned by the rescue center of operation, taking into consideration various factors such as maritime currents, winds, distress location, time of the incident, and the latest information available about the vehicle(IAMSAR, 2022). These probabilities can also be assigned using software like SAROPS (Search and Rescue Optimal Planning System). This software estimates the position of the victims using Bayesian Search Theory and is currently the software used by the U.S. Coast Guard(Kratzke et al., 2010). Once the rescue team has this initial probability map, the path for their aircraft and ships is determined by taking this information into account. One of the advantages of using drones for the search is the fast response and the ability for real-time communication and coordination using data from software without requiring voice communication, as needed in crewed aircraft. The image recognition performed by the drone needs to be fast and reliable. Due to the need to embed the software in a flying vehicle, the weight of processing systems and batteries on the UAV directly impacts the propulsion systems, reducing flight endurance. This restriction limits the processing power available for image recognition, demanding an object detection solution that balances good performance with lightweight hardware.

Training data plays a key role in the performance of a model(Géron, 2017). However, considering that SAR operations prioritize victim care and rescue efforts, it is challenging to acquire data during a real mission. The availability of data for training these systems is a challenge. Despite the existence of some datasets applicable to SAR at sea, the amount of data is generally a factor that can contribute to improving system performance. The use of synthetic data can overcome these limitations, providing theoretically infinite data since it is software-generated. According to (Bird et al., 2020), researchers argue that the use of transfer learning can increase the capability of a model to perform complex tasks after being initially trained on simulations.

In the work developed in (Lima et al., 2023), the training of YOLOv8 was evaluated for different amounts of real and synthetic data using different transfer learning approaches. The strategy of transfer learning used was fine-tuning and freezing layers. From the results, the best model using synthetic data was the one trained first on synthetic data, followed by fine-tuning on real data. It is important to consider that synthetic data can be easily expanded, compared to real-world data, which typically requires significant efforts to acquire. In this context, favoring the use of synthetic data over real data is particularly advantageous.

3 METHODOLOGY

This work will address the question of finding the best balance of real data for fine-tuning a model trained on synthetic data. Different proportions of real and synthetic data will be used for training the model and evaluating the best performance.

3.1 The Model

YOLOv9 is a cutting-edge real-time object detector released in 2024 (Wang and Liao, 2024). It belongs to a series of detection models named YOLO, which stands for "You Only Look Once," first released in (Redmon et al., 2016). The biggest advantage of YOLOv9 is its capability to perform various tasks such as object detection, segmentation, pose estimation, oriented detection, and classification with fast performance even on less powerful systems. The model is available in versions with different numbers of parameters, named with the last letter representing the complexity and consequently the hardware requirements. These versions for object detection are YOLOv9t, YOLOv9s, YOLOv9m, YOLOv9c, and YOLOv9e, respectively representing tiny, small,

medium, large, and extensive sizes. In search areas over the sea, the availability of reliable communication is not common, so a cloud processing approach with high bandwidth demand would not be deployable(Jayalath and Munasinghe, 2021). Considering that drones will need to process all the information internally, YOLOv9t was used considering the performance limitation of embedded systems on UAVs once in a real application the processing power will also be used for different tasks like path-planning, communication management among others.

Different amounts of real-world data will be used during the fine-tuning of the model pre-trained exclusively on synthetic data. The main objective is to find the optimal amount of real data for fine-tuning, which provides the model with proper generalization for detecting survivors in real-life situations.

The total amount of real data available is 10,736. The model will be fine-tuned with 10, 20, 30, 50, 70, and 100 percent of the real data. Each fine-tuning process will be performed directly on the original model trained on synthetic data. The process will not be done sequentially over the same model; for example, a model fine-tuned on 10 percent of data will not be submitted for a new process with 20 percent to create a 30 percent model.

3.2 Datasets and Experimental Setup

3.2.1 SeaDronesSee

SeadronesSee - It is a dataset focused on search and rescue operations using unmanned aerial vehicles in maritime scenarios. Developed by the University of Tübingen, it includes sets of tracks for object detection, single-object tracking, and multi-object tracking (Varga et al., 2021). The images also come with metadata such as altitude, camera angle, field of view, etc. The dataset used in this work is Object Detection v2, which contains a total of 10,477 images. These images are distributed among training, validation, and testing sets, aimed at an evaluation leaderboard managed by the researchers who developed the dataset. The annotation format is compatible with the COCO dataset (Lin et al., 2015), a widely used image dataset.

3.2.2 Synthetic

Synthetic – This dataset was generated using Unreal Engine 4 (Games,) with the Environment Project add-on (DotCam et al., 2022). It features a customizable sea environment where characteristics such as wave direction, size, and frequency can be easily modified. Additionally, it includes configurable sky settings and buoyancy configurations for objects and

Figure 1: Image from SeaDronesSee.

characters on the sea. Utilizing these features, a distress sea scenario was created and automatically annotated, producing images and bounding boxes used as ground truth for object detection models. Developed by undergraduate students from the University of South-Eastern Norway (Pettersvold et al., 2023), the dataset contains 9,500 labeled images. These images are captured in a virtual environment by a UAV flying over characters, objects, and boats on the sea. The virtual environment is set with a clear sky, calm sea, and good weather, mirroring the scenario of the SeadronesSee dataset. Annotations are formatted in YOLOv5 PyTorch TXT (Ultralytics, 2022).

Figure 2: Synthetic image acquired on Unreal 5 simulation.

3.3 Training, Validation and Testing Split

Initially, the annotations of the real-world dataset were converted to the format YOLOv5 PyTorch. Three types of models were generated. The first type is a model trained with all the synthetic data available. The second type are models trained with just real data, in different quantities as will be explained further below. And the third type are models trained with all synthetic data and afterwards refined using transfer

learning through fine-tuning technique over the same quantities of real world data used for the second type.

The initial validation during the training process of the synthetic data will be performed in the real data aiming to guide the system to generalize over the target data. During the fine-tuning process, the validation will be done over a different subset of images from the real dataset. Therefore, the real-world dataset images were divided into 4 subsets as shown in Figure 3.

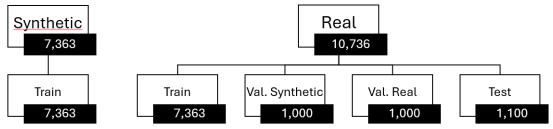

Figure 3: Dataset division.

Where in synthetic data:

- Train - Data for training during the pre-training process.

And the real data division is done as:

- Train - Data used on training or fine-tuning process.
- Val. Synthetic - Data for validation during the pre-training with synthetic dataset.
- Val. Real - Data for validation during the training or fine-tuning process.
- Test - Data used for the final test phase for comparing the models.

The training and validation amounts of data were kept the same for the synthetic and the real datasets, allowing for a good comparison and an easy understanding of the mixed proportions used in the tests. Therefore, a proportion of 10% of real data in a transfer learning model indicates that 736 real images were used for fine-tuning a model pre-trained on 7,363 synthetic images. The optimizer was set to automatic, with a learning rate and momentum set by the YOLOv9 algorithm. The training process and the fine-tuning process were both set to 80 epochs. The starting weights for the model were randomly assigned. To maintain reproducibility, a seed of 42 was chosen for training and fine-tuning the model.

4 RESULTS

The models were evaluated using mAP (mean average precision)(Everingham et al., 2014) as the main metric. It was chosen because it balances precision and recall, considering both false positives and false negatives, and also because it is the benchmark metric used for computer vision(V7Labs, 2022). This metric calculates the accuracy for object detection systems measuring the precision and recall based on the bounding boxes predicted by the model and the bounding boxes provided as ground-truth.

The mAP 50-95, commonly used as the benchmark in computer vision models, measures the mean average precision for IoU (intersection over union) of bounding box overlap thresholds ranging from 0.5 to 0.95. Since the main concern of this research is to detect the presence of survivors in the sea , rather than their exact position within the image, bounding box precision is not as important as detection accuracy. A high overlap requirement may result in more false negatives, which is critical to avoid in search and rescue efforts(Qingqing et al., 2020). Therefore, the average precision at an IoU threshold of 0.5 (mAP50) will be used as the primary metric in this work.

To facilitate the visualization due to the number of lines, the training graphs were divided into two. These data shows the mAP50 considering the training using images of Human and Boat. Figure 4 represents the graphs for the training of 10%, 20%, 30% and pure synthetic while the Figure 5 represents the graphs for the training of 50%, 70%, 100% and pure synthetic. The same color was used for each percentage with the dark one representing the fine-tuning using synthetic data and the light one representing the model trained with the real data from the scratch.

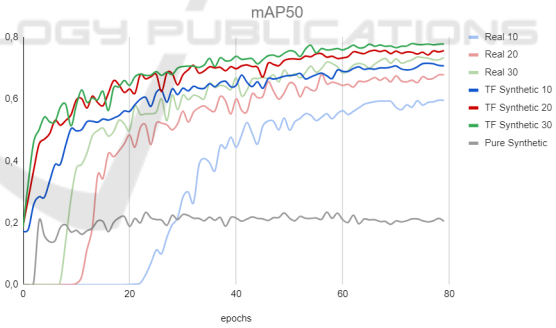

Figure 4: Training curve for 10%, 20%, 30% and synthetic.

In Table 1, the performance of the model for both classes Boat and Human(named as ”All”) is shown for different amounts of real data used for training and for transfer learning. The column ”Real” indicates the performance when using only real data, while the column ”TF Synth” indicates the values when using a specific percentage of real data for transfer learning after the training was completed with the whole amount of synthetic data. The column named ”Increment” indicates the improvement in the model’s detection performance achieved by using synthetic data compared to the use of just real data.

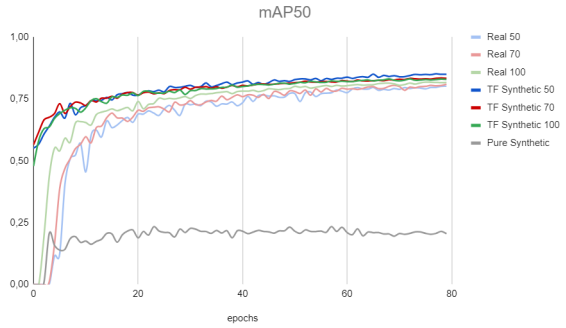

Figure 5: Training curve for 50%, 70%, 100% and synthetic.

Table 1: Models performance for all classes.

mAP 50 for all classes			
% of real data	Real	TF Synth	Increment
10%	0.460	0.523	13.70 %
20%	0.586	0.618	5.46 %
30%	0.653	0.693	6.13 %
50%	0.760	0.800	5.26 %
70%	0.795	0.831	4.53 %
100%	0.820	0.831	1.34 %

The results from Table 1 can be seen on the Figure 6. The yellow line at the bottom indicates the increment values for different amounts of real data.

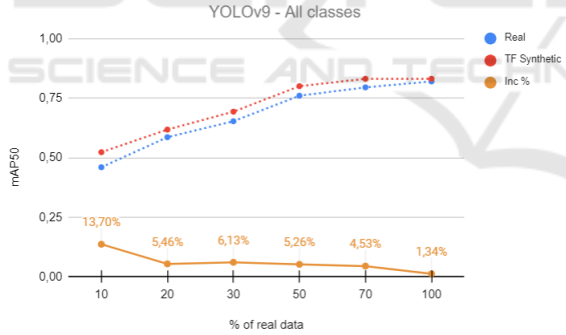

Figure 6: Performance for all classes.

In Table 2 the performance of the model detecting the class human is shown. The correspondent graphic representation of this data is shown in Figure 7.

The performance achieved on boat detection can be seen in Table 3. The correspondent graphic representation of this data is shown in Figure 8.

5 DISCUSSION

The training curves in Figure 4 and 5 reveals a trend: a pre-training model on synthetic data converges the mAP50 faster and keeps a higher performance during all epochs when exposed to the same proportion of real data.

Table 1 and the graph in Figure 6 show a significant increase in performance for all classes when synthetic data is used for pre-training the model compared to just real data for all the different amounts

Table 2: Models performance for class Human.

mAP 50 for Human			
% of real data	Real	TF Synth	Increment
10 %	0,329	0,371	12,76 %
20 %	0,503	0,546	8,54 %
30 %	0,595	0,615	3,36 %
50 %	0,644	0,69	7,14 %
70 %	0,681	0,734	7,78 %
100 %	0,718	0,728	1,39 %

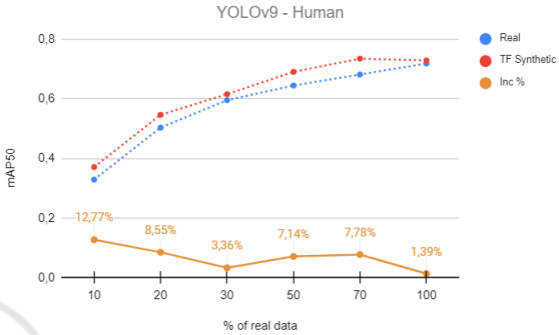

Figure 7: Performance for class Human.

Table 3: Models performance for class Boat.

mAP 50 for boats			
% of real data	Real	TF Synth	Increment
10 %	0,590	0,675	14,4 %
20 %	0,669	0,691	3,28 %
30 %	0,711	0,771	8,43 %
50 %	0,876	0,876	0 %
70 %	0,909	0,928	2,09 %
100 %	0,922	0,934	1,3 %

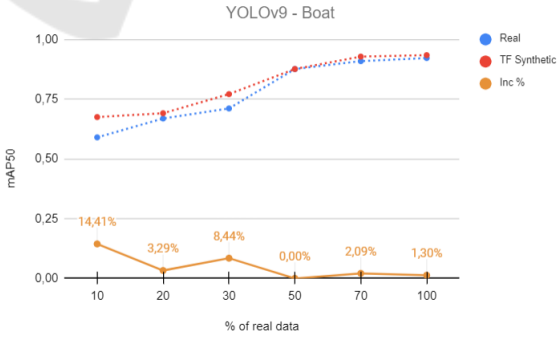

Figure 8: Performance for class Boat.

of real images used. The progressive smaller increments as a bigger proportion of real data is used can be expected since the availability of more real data naturally makes the model perform better.

The same trend of best performance can be seen on Table 2 and graph of the Figure 7, which is the main focus for recognition on SAR missions. On the other hand, the use of synthetic data shows a smaller contribution on the case of boat recognition as can be seen in Table 3 and graph of the Figure 8.

A remarkable point to mention is that according to Table 1, when 70% of real data was used for the transfer learning with the synthetic data, the model performed better (0.831) than using 100% of just the real data(0.820). This behaviour can also be observed for Humans on Table 2 where 70% of real data in transfer learning (0.734) performed better than 100% of real data on training (0.718). Finally for the boat recognition occurs the same trend when transfer learning with 70% (0.928) has a better performance than 100% of real data (0.922) as can be seen on Table 3.

Table 4 shows the model performance for the training with only synthetic data as "Synthetic", only 10% of real data as "Real" and for the transfer learning with 10% of real data on the model pre-trained on synthetic images as "TF 10%". The first 3 columns indicates the classes and the last columns indicates the amount of real and synthetic data used in each process.

Table 4: Performance for using 10% of real data.

Training strategy	All	Human	Boat	Synth img	Real img
Synthetic	0,167	0,012	0,322	7,370	0
Real 10%	0,460	0,329	0,590	0	737
TF 10%	0,523	0,371	0,675	7,370	737

This table shows that the performance of the model when trained with just synthetic data is very low, even if the amount of data is high (7,370). But when a small quantity of real data (10% = 737) is used for the transfer learning process, the model presents a better performance than with only 10% of real data, resulting in a 13.7% performance boost, achieving an mAP50 of 52.3%. These results shows that even for a small quantity of real data, the addition of synthetic data brings a improvement on the model detection capability. This is particularly interesting once synthetic data can be generated rapidly and at large scale.

The improved performance when combining synthetic and real data can be related to a higher amount of data provided by the synthetic images, a larger diversity can also lead a better capacity for the model to generalize. The data provided on the virtual scenario can also provided different elements not present on real-world dataset, helping the model in pattern recognition. However, the similarity between both type of data domains can also be a limitation when the virtual world can not accurately mimic real-world conditions. Another challenge is when the opposite effect can happen with the lack of real datasets containing variations that can be easily generated on virtual world (e.g., extreme weather or lighting conditions).

The results can be compared with the findings of (Krump and Stütz, 2020), which investigated vehicle detection performance using images acquired by UAVs. The evaluation compared models trained on real, synthetic, and mixed datasets using YOLOv3. The reality gap was identified as the main factor impacting detection accuracy, leading to the conclusion that combining real and synthetic data improves performance. While their work explored how context, environmental conditions, and simulation parameters influence detection accuracy, and the current work focuses on transfer learning improvements in a maritime environment, the current study confirms the assumptions of the previous work by demonstrating that incorporating synthetic data with real data during the training phase leads to performance improvements, corroborating the results achieved by (Krump and Stütz, 2020).

6 CONCLUSION AND FUTURE WORKS

With the progressive use of UAVs in different fields and the emergence of powerful and resource-efficient artificial intelligence models, the demand for data has increased. In this work, two different datasets were used: a real-world dataset specifically designed for aerial recognition using drones in a maritime environment, and a synthetic dataset developed in Unreal Engine 4 that reflects a similar scenario. The model YOLOv9t was trained under different configurations to evaluate the impact of transfer learning with real and synthetic data. Initially, the model was trained from scratch using different amounts of real data. Subsequently, other models were pre-trained on the full set of synthetic data and then fine-tuned using the same amounts of real data as the initial models. Finally, the performance of models trained with the same amount of real data was compared, with and without fine-tuning on synthetic data. The results showed a better object detection score when using the transfer learning process, with models pre-trained on synthetic data performing better than those trained exclusively on real data. This study demonstrates that

using synthetic data to train object detection systems with YOLOv9t is a valid approach for overcoming the challenges of real-world data acquisition. The obtained results show that synthetic data is a feasible and effective tool, particularly in the context of search and rescue operations using transfer learning methods. The performance improvement when exposing the model to even 10% of real data is notable. Special attention should be given to the observation that transfer learning with 70% of real data performed better than models trained on 100% real data. This approach of using small amounts of real data opens up the possibility of training models even when real-world data is sparse, as synthetic data can be generated rapidly and in large quantities.

The capability of easy virtual dataset generation can be explored to address the creation of a massive amount of synthetic data compared to real data. Higher similarity between synthetic and real image datasets can also be considered to improve the model, or studies could focus on increasing the diversity of synthetic datasets to achieve better generalization for real-world recognition. Further work can be done by expanding the datasets to include different weather, lighting, and sea conditions for both real and synthetic data. The expansion of the evaluation to different domains, such as terrestrial SAR operations, can also be explored. Incorporating different noise sources, like dust and humidity affecting camera lenses, can further simulate real-world conditions. According to (Krump and Stütz, 2021), the main difference between real and synthetic data, referred to as the "reality gap," is related to general coloration, the absence of noise, and the lack of fine structures. This opens the possibility for further research to bridge this gap.

Implementation and testing in real-world scenarios can be explored, evaluating the integration of all solutions with hardware constraints and associated challenges. These constraints may include factors such as different camera resolutions, embedded processing power, and image stabilization systems (gimbal). Hardware limitations could significantly impact performance, and comparisons of the current model YOLOv9t with different architectures can help optimize factors such as recognition time, training requirements, and effectiveness.

REFERENCES

Bird, J. J., Faria, D. R., Ekárt, A., and Ayrosa, P. P. S. (2020). From simulation to reality: Cnn transfer learning for scene classification. In *2020 IEEE 10th International Conference on Intelligent Systems (IS)*, pages 619–625.

Dabbiru, L., Goodin, C., Carruth, D., and Boone, J. (2023). Object detection in synthetic aerial imagery using deep learning. In Dudzik, M. C., Jameson, S. M., and Axenson, T. J., editors, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 12540 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 1254002.

DotCam, TK-Master, Zoc, and Elble, S. (2022). Environmentproject. https://github.com/UE4-OceanProject/Environment-Project.

Everingham, M., Eslami, S. M. A., Gool, L. V., Williams, C. K. I., Winn, J. M., and Zisserman, A. (2014). The pascal visual object classes challenge: A retrospective. *International Journal of Computer Vision*, 111:98 – 136.

Games, E. Unreal engine. https://www.unrealengine.com. Accessed: 2024-02-22.

Géron, A. (2017). *Hands-on machine learning with Scikit-Learn and TensorFlow : concepts, tools, and techniques to build intelligent systems*. O'Reilly Media, Sebastopol, CA.

IAMSAR, I. (2022). International aeronautical and maritime search and rescue manual. *Mission coordination*, 2.

Jayalath, K. and Munasinghe, R. (2021). Drone-based autonomous human identification for search and rescue missions in real-time. pages 518–523.

Kratzke, T. M., Stone, L. D., and Frost, J. R. (2010). Search and rescue optimal planning system. In *2010 13th International Conference on Information Fusion*, pages 1–8.

Krump, M. and Stütz, P. (2020). Uav based vehicle detection with synthetic training: Identification of performance factors using image descriptors and machine learning. In *Modelling and Simulation for Autonomous Systems: 7th International Conference, MESAS 2020, Prague, Czech Republic, October 21, 2020, Revised Selected Papers*, page 62–85, Berlin, Heidelberg. Springer-Verlag.

Krump, M. and Stütz, P. (2021). Uav based vehicle detection with synthetic training: Identification of performance factors using image descriptors and machine learning. In Mazal, J., Fagiolini, A., Vasik, P., and Turi, M., editors, *Modelling and Simulation for Autonomous Systems*, pages 62–85, Cham. Springer International Publishing.

Lima, L., Andrade, F., Djenouri, Y., Pfeiffer, C., and Moura, M. (2023). Empowering search and rescue operations with big data technology: A comprehensive study of yolov8 transfer learning for transportation safety. pages 2616–2623.

Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., and Dollár, P. (2015). Microsoft coco: Common objects in context.

Pettersvold, J., Wiulsrod, M., and Hallgreen, S. (2023). Synthetic data generation for search and rescue missions. a novel approach using unreal engine, airsim, and raycast. Bachelor's thesis, University of South-Eastern Norway.

Qingqing, L., Taipalmaa, J., Queralta, J. P., Gia, T. N., Gabbouj, M., Tenhunen, H., Raitoharju, J., and Westerlund, T. (2020). Towards active vision with uavs in marine search and rescue: Analyzing human detection at variable altitudes. In *2020 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, pages 65–70.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection.

Tuśnio, N. and Wróblewski, W. (2022). The efficiency of drones usage for safety and rescue operations in an open area: A case from poland. *Sustainability*, 14(1).

Ultralytics (2022). ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation. https://github.com/ultralytics/yolov5.com. Accessed: 7th May, 2023.

Upton, E. and Halfacree, G. (2014). *Raspberry Pi User Guide*. John Wiley & Sons.

V7Labs (2022). Mean average precision (map) explained: Everything you need to know. Accessed on March 20, 2024.

Varga, L. A., Kiefer, B., Messmer, M., and Zell, A. (2021). Seadronessee: A maritime benchmark for detecting humans in open water.

Wang, C.-Y. and Liao, H.-Y. M. (2024). YOLOv9: Learning what you want to learn using programmable gradient information.

Xing, D. and Tzes, A. (2023). Synthetic aerial dataset for uav detection via text-to-image diffusion models. In *2023 IEEE Conference on Artificial Intelligence (CAI)*, pages 51–52.

SCITEPRESS
SCIENCE AND TECHNOLOGY PUBLICATIONS