



Enhancing Detection of Remotely-Sensed Floating Objects via Data Augmentation for Maritime SAR

Haitham Mahmoud¹ · Ibnu F. Kurniawan^{3,4} · Adel Aneiba¹ · A. Taufiq Asyhari² 

Received: 23 July 2023 / Accepted: 12 April 2024
© The Author(s) 2024

Abstract

A figure of 33,000 search and rescue (SAR) incidents were responded to by the UK's HM Coastguard in 2020, and over 1322 rescue missions were conducted by SAR helicopters during that year. Combined with Unmanned Aerial Vehicles (UAVs), artificial intelligence, and computer vision, SAR operations can be revolutionized through enabling rescuers to expand ground coverage with improved detection accuracy whilst reducing costs and personal injury risks. However, detecting small objects is one of the significant challenges associated with using computer vision on UAVs. Several approaches have been proposed for improving small object detection, including data augmentation techniques like replication and variation of image sizes, but their suitability for SAR application characteristics remains questionable. To address these issues, this paper evaluates four float detection algorithms against the baseline and augmented datasets to improve float detection for maritime SAR. Results demonstrated that YOLOv8 and YOLOv5 outperformed the others in which F1 scores ranged from 82.9 to 95.3%, with an enhancement range of 0.1–29.2%. These models were both of low complexity and capable of real-time response.

Keywords Remote sensing · Maritime SAR · Data augmentation · Float detection

Introduction

In 2020, the UK's HM Coastguard responded to 33,000 incidents for search and rescue (SAR) missions, conducting over 1322 SAR helicopter missions between April and September (Maritime and coastguard agency, GOV.UK, 2022; Bullock, 2021). The SAR helicopters completed 1226 rescues and

232 assistance missions in the year ending March 2021, with 46% being rescues or recovery missions (Alemohammad et al., 2020). In the US, 19,951 SAR missions were conducted by the Coast Guard in 2017, resulting in 618 lives lost (Baetens et al., 2019). A total of 16,573 cases of SAR saved 3417 lives in one year (Baetens et al., 2019). Finding missing people and items is one of these operations' most challenging and costly aspects. Hence, UAVs have been used to aid SAR missions. Helicopters and drones now use an array of tools, including cameras, reflectors, transponders, tracking systems, and detectors for mobile phones. SAR operations are also revolutionized by using artificial intelligence and computer vision to process images captured by UAVs.

These technologies enable rescuers to cover more ground in less time with greater accuracy, enhancing the chances of locating missing individuals and items. Moreover, deploying these tools minimizes the risk of human injury in challenging terrains or hazardous environments. A drone or other unmanned vehicle allows rescuers to cover vast areas without risking their lives. Ongoing technological advancements continually improve these tools, offering exciting possibilities for saving more lives and reducing the costs associated with search and rescue missions.

✉ A. Taufiq Asyhari
taufiq-a@ieee.org

Haitham Mahmoud
haitham.mahmoud@bcu.ac.uk

Ibnu F. Kurniawan
ibnufebry@unesa.ac.id

Adel Aneiba
adel.aneiba@bcu.ac.uk

¹ School of Computing and Digital Technology, Birmingham City University, Birmingham B4 7XG, UK

² Data Science Department, Monash University, Indonesia, Tangerang, Banten 15345, Indonesia

³ Centre for Computational Science and Mathematical Modelling, Coventry University, Coventry CV1 2TL, UK

⁴ Department of Data Science, Universitas Negeri Surabaya, Surabaya 60231, Indonesia

Utilizing computer vision on unmanned vehicles poses challenges, specifically in small object detection on the surface area, defined as regions with dimensions ≤ 32 by 32 pixels (Lin et al., 2014). While current state-of-the-art models excel in detecting medium and large-sized objects, detecting small objects remains challenging due to limited information. This makes it hard to distinguish from the background or other categories. Deep architecture, with sequential pooling layers, attenuates small areas to reduce noise-induced distortions. However, this may lead to imprecise object locations, especially when the object appears in multiple places. Precision requirements for small object detection are essential.

Several approaches aim to improve small object detection. One method, presented in Kisantal et al. (2019), enhances data by copying and pasting small objects multiple times. This increases the number of images in the dataset with small objects, enhancing alignment between predicted anchors and small ground-truth objects. The approach reports a relative improvement of over 7% during the object detection phase. This technique effectively addresses the challenge of detecting small objects.

This paper presents a modified network training approach involving images with differing scales and comparing a ground truth image with objects within a specified range (Singh & Davis, 2018). By addressing the challenge of learning objects of radically different scales together, this technique significantly reduces computational complexity (Singh et al., 2018). Therefore, this paper improves the performance of float detection using the suggested data augmentation algorithm for maritime SAR. In the case of minority classes, the data augmentation produces artificial samples that assist in detecting them. For this work, we utilized four float detection algorithms. Finally, we evaluated and compared the performance of aerial datasets' baseline version with augmented ones. We used EfficientDet, Faster R-CNN, YOLOv5, and YOLOv8 algorithms. Faster R-CNN is a widely used algorithm; meanwhile, EfficientDet excels in accuracy and efficiency, and recent studies highlight the real-time high detection accuracy potential of YOLOv5 and YOLOv8 algorithms. This evaluation compares the baseline dataset and augmented data, demonstrating the algorithm's effectiveness. Furthermore, we assessed the effectiveness of our approach through comparisons with two benchmark studies.

This paper is organized as follows: Sect. 2 examines the existing studies on similar applications and existing datasets. Section 3 describes the architecture, data augmentation and remote detection algorithms, dataset and evaluation metrics. Section 4 discusses the results. Finally, Sect. 5 summarizes the findings of the study and addresses the future work.

Related Works

Datasets containing aerial images are primarily considered in this application. These aerial images are usually acquired using UAVs equipped with standard cameras or satellite imaging systems. Due to these varied capture methods, there is a significant difference in the spatial resolution of the collected images. Satellite cameras can have resolutions as low as 30 m per pixel, while drone-mounted cameras typically yield 1–10 cm per pixel (Burke et al., 2019). It may be challenging to train machine learning models on aerial image datasets due to the discrepancy in spatial resolution. Some classes may have fewer high-resolution images available than others, which may cause class imbalances. Consequently, some classes can have very few instances for the training process. The model generated from such a dataset will have a reduced ability to learn from underrepresented classes, resulting in reduced accuracy and performance (Krawczyk, 2016; Das et al., 2022).

Many satellite datasets have a low spatial resolution, which is typically used to classify scenes (e.g. industrial, residential, river, forest, crop, etc.) such as Eurostat (Helber et al., 2019, 2018) BigEarthNet (Sumbul et al., 2019). Moreover, the SARSAT system has been used to detect emergency beacons on individuals, vessels, and aircraft (Comps, 1984). Another approach to SAR involves collecting and analysing environmental and oceanographic data (Futch & Allen, 2019). Furthermore, some studies have used datasets of past SAR incidents to analyse the spatial distribution and frequency of such incidents called SISAR (Stoddard & Pelot, 2020). SISAR data covers all of Canada's coastal SAR areas. There are approximately 36,000 incidents in the dataset, each of which has a georeferenced position associated with it. Visualizing the spatial distribution of incidents can be accomplished by plotting the location of incidents on the analysis map. Two large datasets are available for detecting objects in satellite photographs: xView (Lam et al., 2018) and DOTA (Xia et al., 2018), which have spatial resolutions of 20–40 cm. Both cases involve passenger cars as the smallest marked objects. XView includes more than one million instances of artificial objects from 60 categories, all taken from WorldView-3 satellites at a resolution of 0.3 m per pixel. On the other hand, the DOTA dataset contains 188,282 objects that are grouped into 15 categories. The photos have a wide variety of spatial resolutions, with bounding boxes that are horizontal and oriented. Furthermore, the Ship Detection Challenge Dataset is an application-specific dataset derived from satellite images (Helber et al., 2019). Airbus developed the system to help SAR services locate missing ships from an aerial perspective. The dataset consists of 131 thousand instances of ships annotated on satellite images with a resolution of 1.5 ms.

On the other hand, Medium spatial resolution data (1–10 m) captured by UAVs are used for more challenging semantic segmentation tasks. For example, the LandCoverNet (Alemohammad et al., 2020) contains seven land classifications (e.g., water, snow, woody), LandCover.ai (Boguszewski et al., 2021) with three types of land allocation (building, woodland, and water), and 95-Cloud ALCD Reference Cloud Masks aimed at semantic segmentation of clouds (Mohajerani & Saeedi, 2019; Mohajerani et al., 2018; Mohajerani & Saeedi, 2020; Baetens et al., 2019).

In addition to the higher spatial resolution of drone datasets, the primary difference between them is the variation in camera orientation, i.e. the pitch angle of the camera. Images from satellites are always oriented vertically in datasets containing satellite images. The most closely related aerial datasets to the SAR, which are publicly available, are listed as shown in Table 1. Most of the data acquired by drones is devoted to detecting and tracking people and cars. A dataset containing 10,209 images and 263 video clips containing almost 180 thousand frames is called Vis-Drone (Zhu et al., 2020). More than 235,000 objects are identified in this dataset, primarily consisting of urban images in which most of which are taken vertically and with low-angle, oblique camera orientation. The objects are grouped into ten classes (e.g. pedestrian, car, bus, truck, bicycle). Furthermore, the Stanford Drone Dataset (Robicquet et al., 2016) consists of over 920 thousand video frames with 185281 labelled objects categorized into six classes (pedestrian, bicycle, skateboard, cart, car, bus). This tool was developed primarily for tracking and trajectory forecasting in urban environments. Additionally, drone datasets are frequently produced for a specific purpose (Bonet et al., 2020; Gasienica-Jozkoway et al., 2021; Jiang & Zhang, 2020; Liu et al., 2020; Ren et al., 2015). Because the paper only concerns SAR missing persons, it

focuses on utilizing the AFO (Gąsienica-Józkowy et al., 2021).

As previously mentioned, training machine learning models on aerial image datasets can be challenging due to spatial resolution discrepancies. Some classes may have fewer high-resolution images available than others, which may cause class imbalances, and some classes may have very few instances for training. Therefore, a data augmentation method is discussed in this study to address this issue effectively by generating synthetic images with desirable characteristics, such as rotation, flip, noise, contrast, and brightness, for underrepresented classes. The augmented images can increase the instances of these classes in the dataset, improving the balancing of classes and providing the model with more training data. The model may be trained more robustly and correctly on aerial image datasets by selectively applying data augmentation to under-represented classes while keeping instances of well-represented classes.

Methodology

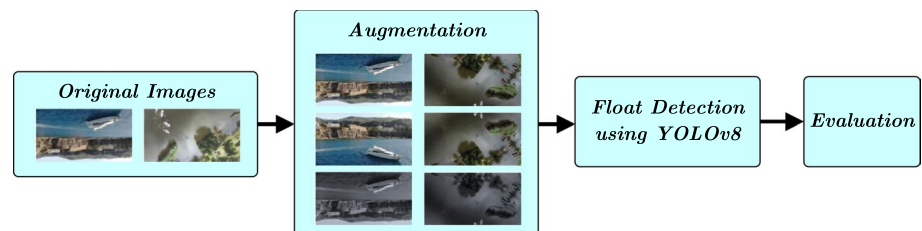
Architecture

Our proposed processing pipeline (see, Fig. 1) consists of four key components, such as data storage, data augmentation, float detection algorithms, and performance evaluation. The flow begins with data management which includes data acquisition, data pre-processing, and data storing. This functionality ensures that the data used in the processing is suitable for the purpose, i.e., detecting the floating object. Next, the existing data is enhanced by applying data augmentation techniques, which essentially produce artificial samples of the minority class. Careful operation is required in this step

Table 1 The most related aerial datasets to SAR

Name	Classes	Size	Application
AFO (Gąsienica-Józkowy et al., 2021)	6	3647 images	Maritime SAR (i.e., humans, boats, kayak, etc.)
Vis-Drone (Zhu et al., 2020)	10	10,209 images	Urban detection (i.e., pedestrian, car, bus, etc.)
Stanford Drone (Robicquet et al., 2016)	6	920 videos	Mobility prediction (i.e., pedestrian, bicycle, etc.)

Fig. 1 Processing flow for maritime SAR detection



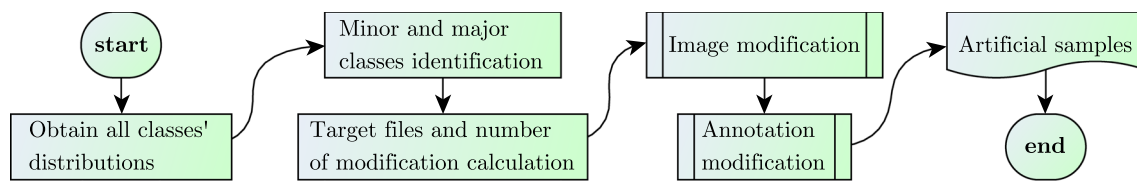


Fig. 3 Flowchart of the augmentation process

number of most-represented classes too much. It also aims at achieving the final ρ efficiently. Finally, the engineering process is executed for the targeted images and their annotations. The image data are modified using the Albumentation library (Buslaev et al., 2020). The runtime complexity for such a process is $\Theta(m)$, where m is the number of candidate files. The Θ notation refers to the asymptotic value, which indicates a tight bound for a runtime complexity (Singh et al., 2018).

Remote Float Detection

An evaluation of four different object detection algorithms in an image was conducted in this study. We used EfficientDet, Faster R-CNN, YOLOv5, and YOLOv8 algorithms. Faster R-CNN is a popular and widely used algorithm in object detection, while EfficientDet is known for its high accuracy and efficiency. Several recent studies have demonstrated the potential of YOLOv5 and YOLOv8 algorithms. Below is a brief description of each algorithm:

EfficientDet (Tan et al., 2020) utilizes a compound scaling method is employed to achieve high accuracy as well as efficiency. Bidirectional feature pyramid networks (BiFPNs) fuse features across scales to improve accuracy and robustness. Aside from that, it has been demonstrated that it is more accurate and faster than other object detection models (Chen et al., 2021).

Faster R-CNN (Ren et al., 2015) utilizes a two-stage detection process. A Region Proposal Network (RPN) generates object proposals, while a classification network is used to classify the proposals. The RPN generates a set of object proposals by analyzing the input image's spatial positions and scales. Using each proposal as input, the classification network produces a class label and bounding box coordinates for the objects in the proposal.

You Only Look Once (YOLO) (Redmon et al., 2016) uses a neural network to predict bounding boxes for each object in an input image while simultaneously predicting class probabilities for each object. Two versions of YOLO are used in this study, YOLOv5 and YOLOv8, each with its own strengths and weaknesses. The YOLOv5 version is simpler, while the YOLOv8 version is more accurate and faster. The accuracy is expected to be enhanced by 3.2–9.3% depending on the dataset size (AugmentedStartups, 2023).

Dataset

As previously mentioned, the AFO (Aerial dataset of Floating Objects) (Gąsienica-Józkowy et al., 2021) is a publicly available dataset containing images of floating objects. The dataset can be useful for developing an object detector system for beach lifeguards or evacuation purposes. There are six classes in this dataset, each of which represents possible objects found on the beach, such as human (class id 1), wind/sup-board (2), boat (3), buoy (4), sailboat (5), and kayak (6). The default training set configuration provided by its author contains more than 6000 class instances and is dominated by the human class. In terms of the imbalance ratio ρ , defined in (Buda et al., 2018), this dataset has the $1/\rho \leq 0.1$.

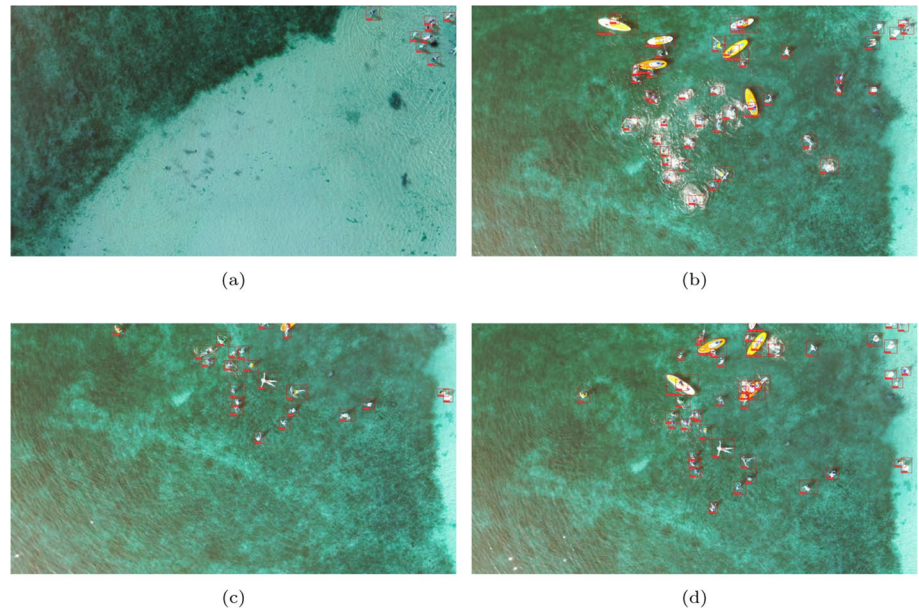
After the augmentation process, the new dataset has better distribution across classes, thus reducing the imbalance ratio. A visual representation of the dataset through samples in Fig. 4 provides insights into images with different compression percentages. Although the size of the dataset may be reduced by using compression techniques, this information is essential for detection tasks. The key features and characteristics of small objects can be preserved when compression methods are carefully selected, enhancing or maintaining the detection accuracy of small objects.

Evaluation Metrics

Several performance metrics are used to evaluate the methodologies used, including; Precision, Recall, F1score, mAP50 and mAP50-90, as follows. Precision refers to the proportion of objects detected by the correct model, e.g., the model's ability to accurately detect vessels, buoys, or life rafts relevant to SAR missions. This metric is defined as

$$\text{Precision} = \frac{TP}{TP + FP},$$

where TP and FP are the rates of True Positives and False Positives, respectively. Recall measures the model's capability to identify all relevant objects within the SAR image that may require search and rescue efforts, that is, the model's ability to detect all actual objects within the SAR image. This metric is denoted by

Fig. 4 Samples of the dataset

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}},$$

where FN describes the number of False Negatives. An F1 score measures the model's performance in detecting objects relevant to the SAR mission in aerial images by combining Precision and Recall. This score is given by

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}.$$

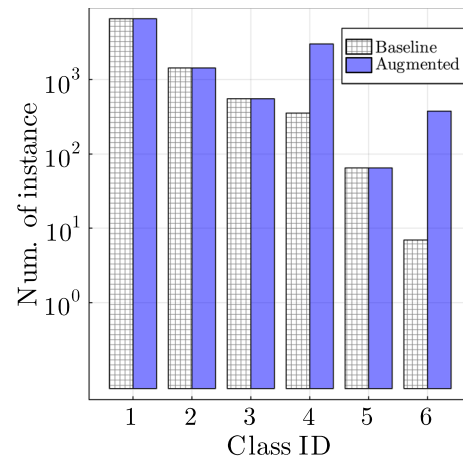
The final two metrics, mAP50 and mAP50-95, evaluate how successful the model is at identifying objects relevant to SAR missions in aerial images, even when they are partly obscured or overlapped. A model's ability to detect relevant objects is assessed using Intersection over Union (IoU) thresholds. This metric is defined as

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^N \text{AP}_i.$$

A comparison of all the above-mentioned metrics obtained from the training process utilising baseline and augmented datasets is performed. This step is applied to measure the enhancement level made by these datasets.

Results and Discussion

As seen in Fig. 5, the sample augmentation algorithm maintains the number of the two most-represented classes' instances while increasing the least-represented ones. The kayak instances' have increased significantly by two magnitudes. This is obtained by setting the final $1/\rho$ to 0.5. This

**Fig. 5** Comparison of two dataset versions' class distribution. The y-axis is in the log scale

indicates the sample augmentation process has been applied carefully to prevent an excessive increase in instances for well-represented classes. A careful balance of class instances through augmentation is necessary to ensure that the dataset continues to reflect the real-world distribution of objects. Taking into account the class imbalance issue in aerial detection. Having fewer instances of some classes leads to biased model performance and decreased accuracy for underrepresented classes. It is intended that the algorithm will provide the model with a wider variety of examples of the least-represented classes, which will enhance its ability to detect and classify those classes in real-life situations.

To evaluate performance for baselines and augmented datasets, we have used EfficientDet, Faster C-RNN, YOLOv5 and YOLOv8, as shown in Table 2. We provided a

Table 2 Performance evaluation of the Results using EfficientDet, Faster C-RNN, YOLOv5 and YOLOv8 for baseline and augmented data

Algorithm	Class	Recall		Precision		F1		mAP50		mAP50-90	
		B	A	B	A	B	A	B	A	B	A
EfficientDet	A	7.2	28.4	73.8	83	13.12	42.32	4.7	30.2	1.5	12.4
	0	3.1	23.3	12.6	53.5	4.95	32.46	1.9	27.5	3.9	7.24
	1	0	0	100	100	0	0	0	2.10	0	3.38
	2	0	0	100	70.6	0	74.56	4	72.8	1.5	30.6
	3	0	0	100	100	0	0	0	4	0	1.3
	4	0	0	100	100	0	0	0	2.3	0	5
Faster C-RNN	5	43.40	68.3	30.40	73.8	35.76	70.94	26.3	72.2	8.6	35.7
	A	21.6	33.6	85.6	90.5	36.5	52.1	25.6	53.3	11	27.2
	0	37.8	27.7	59.1	66.2	42.5	31.4	45.3	35.3	13.5	9.5
	1	0	0	100	100	0	0	0	0.51	0	0.1
	2	0	85.4	71.6	100	0	91.8	7.46	85.9	2.3	38.8
	3	0	0	100	100	0	0	1.1	3.85	0.32	0.8
YOLOv5	4	0	0	100	100	0	0	6.3	99.5	2.58	59.7
	5	91.8	88.3	76.8	83.2	83.4	82.5	93.6	94.8	47.5	45.3
	A	95.4	96.2	93.9	94.5	94.64	95.3	95.3	95.8	57.1	57.2
	0	89.1	90.1	94.1	93.9	91.53	91.9	92.7	92.7	41.8	41.2
	1	0	00	0	0	0	0	0	0	0	0
	2	100	100	90.5	92.8	95.01	96.2	99.5	99.5	79.4	79.5
YOLOv8	3	92.5	94.6	92.1	92.3	92.29	93.4	89.4	91.6	33.2	33.1
	4	0	0	0	0	0	0	0	0	0	0
	5	100	100	98.9	99.1	99.44	99.5	99.5	99.5	73.8	75.1
	A	75.3	92	92.4	89	82.9	90.2	78	92	46.4	57.7
	0	84.1	86.6	91.9	84.5	87.8	85.3	90.2	87.4	38	37.7
	1	0	92.8	100	96.5	0	87.1	0	93.6	0	52.3
	2	100	96.3	87	97.5	93	96.8	97.1	95.4	64.1	80.1
	3	68.2	62.8	90.2	86.5	77.6	87.8	83.3	76.4	29.7	22.9
	4	100	100	85.9	70.7	92.4	82.2	98	99.5	75.6	79.6
	5	99.5	99.5	99.5	98.1	99.5	99.2	99.5	99.5	71.2	73.8

Table 3 Mapping of the class names with the notations in the results

Notation	Class name
A	All
0	Human
1	Wind/Sup-board
2	Boat
3	Buoy
4	Sailboat
5	Kayak

comprehensive analysis of the results by showing the results for each class to depict the enhancement of the imbalanced issue in comparison between the baseline and augmented datasets. Table 3 presents a mapping of the names of classes.

The performance is evaluated using Recall, Precision, F1score, mAP50, and mAP50-90. It can be demonstrated that YOLOv8 and YOLOv5 perform better than EfficientDet and Faster C-RNN. This may be due to YOLO's handling of small objects, which is crucial in float detection for

maritime SAR. As previously mentioned, using a feature pyramid network, YOLO detects small objects at multiple scales. YOLO also employs anchor boxes to predict object sizes and locations, allowing it to handle objects of varying sizes better. Faster C-RNN and EfficientDet also use anchor boxes. However, their effectiveness in handling small objects may not be as great as that of YOLO.

Generally, both YOLO algorithms have outperformed each other for different classes. Several factors contribute to this, including the dataset's nature and the objects' characteristics within different classes. In classes with objects of similar sizes and shapes, YOLOv5 may perform better than YOLOv8, while in classes with more complex and diverse objects, YOLOv8 may perform better than YOLOv5. Compared to the Recall for classes 0, 2, 3 and 5, YOLOv5 outperformed them all. In contrast, YOLOv8 successfully detected objects in classes 4 and 1 after augmentation. This shows the ability of both algorithms to detect all relevant objects within each class. Moreover, the Precision of YOLOv5 and YOLOv8 has varied from 90.5 to 99.1

and 70.7 to 100, respectively, for all classes. This measures the model's accuracy by calculating the ratio between correctly detected objects and the total number of objects detected. However, for both baselines and augmented datasets using YOLOv5 and YOLOv8, the F1 score is 94.64%, 95.3%, 82.9% and 90.2%. Furthermore, the performance has been evaluated among a confidence threshold of 50% and an average between 50% and 90% through using mAP50 and mAP50-90. Both algorithms have relatively the same results in the mAP50-90 YOLOv8 outperformed by 0.5%. On the other hand, for mAP80, YOLOv5 has a higher value of 10.7% for all classes.

Depending on the performance metrics, there is a range of enhancement between the two datasets of 0.1% to 29.2%, indicating the efficacy of the augmentation techniques applied as shown in Fig. 6. The results of this study demonstrate the importance of enhancing datasets to improve the performance of computer vision models. A dataset augmentation is expected to yield increased performance improvements as the model complexity and diversity of data increase. YOLOv8 and YOLOv5 are designed to capture a wide range of objects, backgrounds, and lighting conditions and require large and diverse datasets for high accuracy of training. Adding noise or rotating images to the dataset can increase the diversity of the dataset, allowing the model to learn from more varied and representative examples. The generalization and performance of the model can therefore be improved when applied to real-world data, thereby improving its generalization.

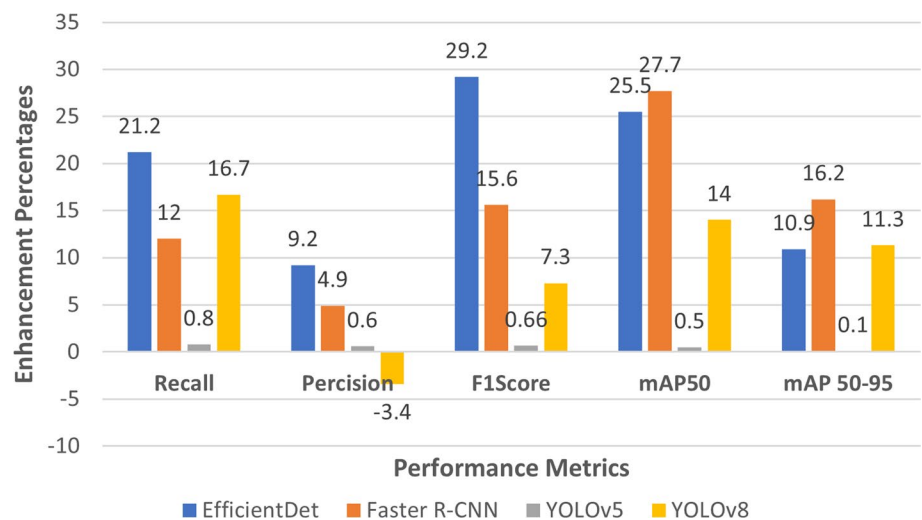
Moreover, both YOLO algorithms can perform real-time detection due to their lower complexity than their counterparts. The complexity of YOLOv8 and YOLOv5 is $O(n^2)$ and $O(n)$, respectively, in which n is the number

of bounding boxes generated per image. This is due to YOLOv8's use of sparse attention, an approach to object detection that reduces the number of computations required. Rather than analysing the entire image simultaneously, sparse attention focuses on only a few spatial locations. Consequently, YOLOv5 and YOLOv8 can support real-time detection. This can be helpful for areas without internet access, but it will not affect the ability to report the presence of any human beings in the area.

Two studies, (Zhu et al., 2023; Wang et al., 2022), have utilized the same dataset. YOLOv5-Augmented leads with the highest MP50 value of 95.8%, indicating superior performance in small object detection as shown in Fig. 7. Following closely is YOLOv8-Augmented with an MP50 value of 92%, demonstrating effective capabilities in SAR scenarios. Other notable performers include YOLOv5 (Wang et al., 2022; Zhu et al., 2023) with an MP50 of 84.5%, and YOLOv7+C2f+SimAM (Wang et al., 2022, 2023) with an MP50 of 83.12%. These evaluations highlight the strengths of specific models in addressing the challenge of detecting small floating objects, with YOLOv5-Augmented leading the pack.

Figure 8 indicates that three classes, namely "Human," "Wind/sup," and "Kayak," exhibit substantial performance improvements in the augmented models compared to the baseline YOLOv5 (Zhu et al., 2023). Specifically, "Human" detection increases from 83.2 to 93.9%, "Wind/sup" goes from 97.7 to 96.5%, and "Kayak" sees a slight improvement from 99.5 to 99.1%. However, it's noteworthy that "Sailboat" detection in the augmented models drops significantly from 91.6 to 70.7%. These variations highlight the nuanced impact of augmentation on different object classes, emphasizing the need for targeted evaluation and improvement strategies.

Fig. 6 Enhancement percentage between baseline and augmented data for the detection of floating objects



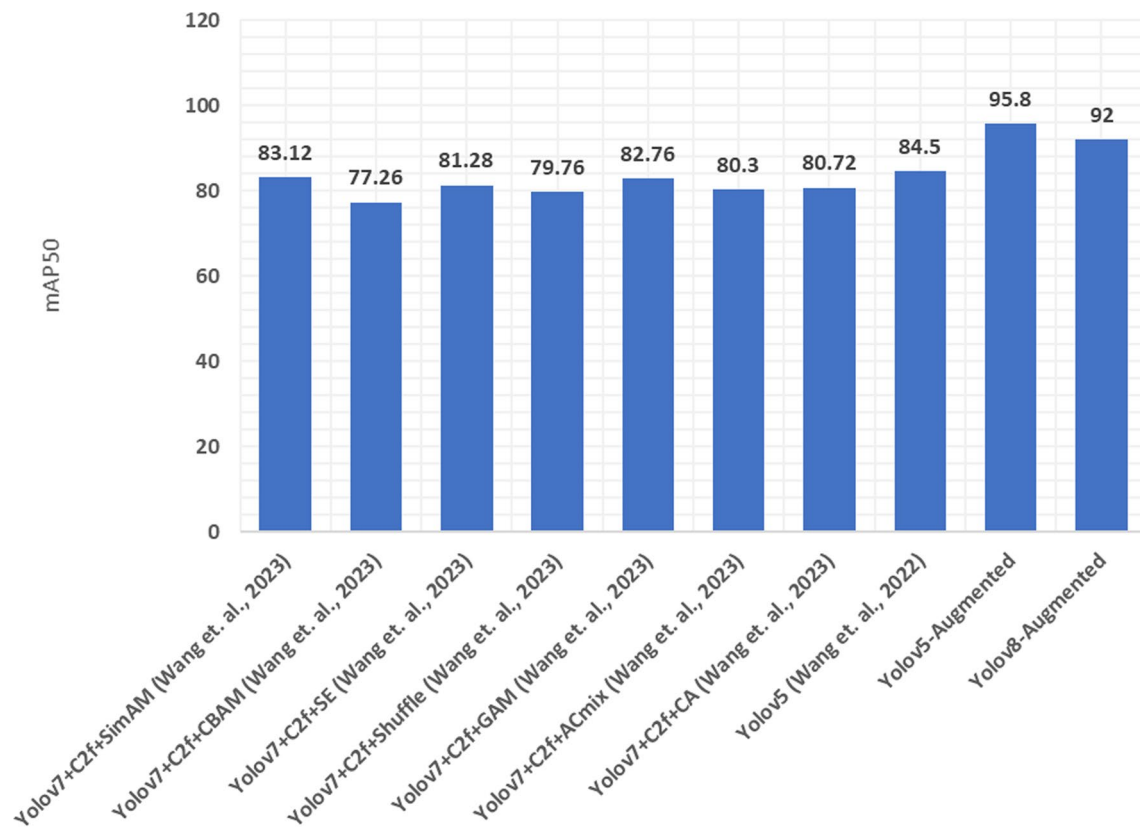


Fig. 7 Performance evaluation with the related studies using mAP50 metric

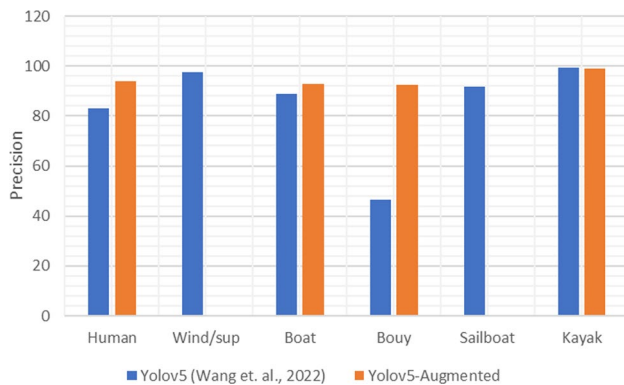


Fig. 8 Performance evaluation with the related studies using precision metric

Conclusion and Future Work

This paper has proposed a data augmentation technique through enhancement procedures, which can reduce overfitting by combining good and noised inputs, but they should be used cautiously to avoid reducing performance. The procedures have included utilizing pixel and spatial-level design to image-annotation pairs, emphasizing

images with under-represented classes while preventing imbalanced data, with the ultimate objective of efficiently achieving the desired dataset characteristics. Four algorithms were evaluated using the maritime SAR dataset with imbalanced issues using Recall, Precision, F1score, mAP50, and mAP50-90. A comparison of YOLOv8 and YOLOv5 with EfficientDet and Faster C-RNN has shown that YOLOv8 and YOLOv5 are more efficient across different performance metrics. For both baselines and augmented datasets, the F1 scores of YOLOv8 and YOLOv5 ranged from 82.9 to 95.3%, with an enhancement range of 0.1–29.2%. These models were both low-complexity and capable of real-time response.

This work can be extended to include the temporal context in the data augmentation (Kim et al., 2020). The models may be able to more accurately depict dynamic changes in the marine environment over time by expanding the current pixel and spatial-level designs to include information from successive frames. This temporal enhancement may be significant in situations where the evolution of characteristics is important, like ship motions.

A second future work is to create an adaptive augmentation technique, which dynamically modifies the degree of enhancement depending on the properties of incoming data (Wang et al., 2021). This should be used to autonomously

identify the best augmentation parameters for various marine situations by utilizing machine learning or reinforcement learning techniques. This tackles the problem of striking a balance between potential overfitting hazards and performance improvements.

Supplementary Information

The dataset used in this study can be accessed from (Gąsienica-Józkowy et al., 2021).

Acknowledgements The work of A. Taufiq Asyhari was started since he was affiliated as a full-time academic staff at Birmingham City University, UK

Author Contributions All authors contributed to the study conception and design. Material preparation and data collection were performed by HM and IFK. Data analysis and evaluation were performed by all authors. Supervision and access to resources were performed by AA and ATA. Funding acquisition was performed by ATA. The first draft of the manuscript was written by HM and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions. This work was supported in part by the British Council COP26 Trilateral Research Initiative Grants for Phase-I and Phase-II. A. Taufiq Asyhari acknowledged support from the Academic Research Startup Grant at Monash University. Ibnu F. Kurniawan acknowledged support from the Directorate General of Higher Education, Research, and Technology, Indonesia.

Declarations

Conflict of interest The authors declared that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- 100, R. (2022). Weed crop aerial Dataset. Roboflow. visited on 2023-04-05. <https://universe.roboflow.com/roboflow-100/weed-crop-aerial>
- Airbus Ship Detection Challenge. <https://www.kaggle.com/c/airbus-ship-detection>
- Alemohammad, S., Ballantyne, A., Bromberg, G., Booth, K., Nakanuku-Diggs, L., & Miglares, A. (2020). Landcovernet: A global land cover classification training dataset.
- AugmentedStartups. (2023). YOLOv8 vs. YOLOv5: Choosing the Best Object Detection Model. AugmentedStartups. Retrieved 13 April 2023, from <https://www.augmentedstartups.com/blog/yolov8-vs-yolov5-choosing-the-best-object-detection-model>
- Baetens, L., Desjardins, C., & Hagolle, O. (2019). Validation of copernicus sentinel-2 cloud masks obtained from maja, sen2cor, and fmask processors using reference cloud masks generated with a supervised active learning procedure. *Remote Sensing*, 11(4), 433.
- Boguszewski, A., Batorski, D., Ziemba-Jankowska, N., Dziedzic, T., & Zambrzycka, A. (2021). Landcover. ai: Dataset for automatic mapping of buildings, woodlands, water and roads from aerial imagery. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 1102–1110).
- Bonet, I., Caraffini, F., Pena, A., Puerta, A., & Gongora, M. (2020). Oil palm detection via deep transfer learning. In *2020 IEEE congress on evolutionary computation (CEC)* (pp. 1–8). IEEE.
- Buda, Mateusz, Maki, Atsuto, & Mazurowski, Maciej A.. (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, 106, 249–259. <https://doi.org/10.1016/j.neunet.2018.07.011>
- Bullock, C. (2021). The UK coastguard publishes 2020 SAR statistics. <https://www.airmedandrescue.com/latest/news/uk-coast-guard-publishes-2020-sar-statistics>
- Burke, C., Rashman, M., Wich, S., Symons, A., Theron, C., & Longmore, S. (2019). Optimizing observing strategies for monitoring animals using drone-mounted thermal infrared cameras. *International Journal of Remote Sensing*, 40(2), 439–467. <https://doi.org/10.1080/01431161.2018.1558372>
- Buslaev, A., Iglovikov, V. I., Khvedchenya, E., Parinov, A., Druzhinin, M., & Kalinin, A. A. (2020). Albumentations: Fast and flexible image augmentations. *Information*, 11(2), 125. <https://doi.org/10.3390/info11020125>. Accessed 2023-11-27.
- Chen, X., Xie, C., Tan, M., Zhang, L., Hsieh, C.-J., & Gong, B. (2021). Robust and accurate object detection via adversarial learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 16622–16631).
- Comps, A. (1984). The search and rescue satellite aided tracking (sarsat) experiment. presentation of cnes processing for the sargos project: Context and orbit processing. In *Satellite motion: Lectures and exercises on space mech* (pp. 749–781).
- Cormen, T. H. *Introduction to algorithms*, 3rd edn. MIT press.
- Das, S., Mullick, S. S., & Zelinka, I. (2022). On supervised class-imbalanced learning: An updated perspective and some key challenges. *IEEE Transactions on Artificial Intelligence*, 3(6), 973–993. <https://doi.org/10.1109/TAI.2022.3160658>. Accessed 2023-07-21.
- Futch, V., & Allen, A. (2019). Search and rescue applications: On the need to improve ocean observing data systems in offshore or remote locations. *Frontiers in Marine Science*, 6, 301.
- Gąsienica-Józkowy, J., Knapik, M., & Cyganek, B. (2021). An ensemble deep learning method with optimized weights for drone-based water rescue and surveillance. *Integrated Computer-Aided Engineering*. <https://doi.org/10.3233/ICA-210649>
- Gąsienica-Józkowy, J., Knapik, M., & Cyganek, B. (2021). An ensemble deep learning method with optimized weights for drone-based water rescue and surveillance. *Integrated Computer-Aided Engineering*, 28(3), 221–235.
- Helber, P., Bischke, B., Dengel, A., & Borth, D. (2018). Introducing eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. In *IGARSS 2018–2018 IEEE international geoscience and remote sensing symposium* (pp. 204–207). IEEE.

- Helber, P., Bischke, B., Dengel, A., & Borth, D. (2019). Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(7), 2217–2226.
- Jiang, S., & Zhang, J. (2020). Real-time crack assessment using deep neural networks with wall-climbing unmanned aerial system. *Computer-Aided Civil and Infrastructure Engineering*, 35(6), 549–564.
- Kim, T., Lee, H., Cho, M., Lee, H. S., Cho, D. H., & Lee, S. (2020). Learning temporally invariant and localizable features via data augmentation for video recognition. In *Computer vision—ECCV 2020 workshops: Glasgow, UK, August 23–28, 2020, proceedings, part II 16* (pp. 386–403). Springer.
- Kisantal, M., Wojna, Z., Murawski, J., Naruniec, J., & Cho, K. (2019). Augmentation for small object detection. *arXiv:1902.07296*
- Krawczyk, B. (2016). Learning from imbalanced data: Open challenges and future directions. *Progress in Artificial Intelligence*, 5(4), 221–232. <https://doi.org/10.1007/s13748-016-0094-0>
- Kurniawan, I. F., He, F., Dunggio, I., Hamidun, M. S., Sirajuddin, Z., Aziz, M., & Asyhari, A. T. (2023). Imbalanced learning of remotely sensed data for bioenergy source identification in a forest in the Wallacea region of Indonesia. *Remote Sensing Letters*, 14(11), 1117–1128. <https://doi.org/10.1080/2150704X.2023.2270107>
- Lam, D., Kuzma, R., McGee, K., Dooley, S., Laielli, M., Klaric, M., Bulatov, Y., & McCord, B. (2018). xview: Objects in context in overhead imagery. *arXiv:1802.07856*
- Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In *Computer vision—ECCV 2014: 13th European conference, Zurich, Switzerland, September 6–12, 2014, proceedings, Part V 13* (pp. 740–755). Springer.
- Liu, Y.-F., Nie, X., Fan, J.-S., & Liu, X.-G. (2020). Image-based crack assessment of bridge piers using unmanned aerial vehicles and three-dimensional scene reconstruction. *Computer-Aided Civil and Infrastructure Engineering*, 35(5), 511–529.
- Maritime and coastguard agency. GOV.UK. (2022). <https://www.gov.uk/government/organisations/maritime-and-coastguard-agency/about>
- Mohajerani, S., & Saeedi, P. (2019). Cloud-net: An end-to-end cloud detection algorithm for landsat 8 imagery. In *IGARSS 2019-2019 IEEE international geoscience and remote sensing symposium* (pp. 1029–1032). IEEE.
- Mohajerani, S., & Saeedi, P. (2020). Cloud-net+: A cloud segmentation cnn for landsat 8 remote sensing imagery optimized with filtered jaccard loss function. *arXiv e-prints* (2001)
- Mohajerani, S., Krammer, T. A., & Saeedi, P. (2018). Cloud detection algorithm for remote sensing images using fully convolutional neural networks. *arXiv:1810.05782*
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*.
- Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., Garnett, R. (Eds.), *Advances in neural information processing systems* (vol. 28). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2015/file/14bfa6bb14875e45bba028a21ed38046-Paper.pdf
- Robicquet, A., Sadeghian, A., Alahi, A., & Savarese, S. (2016). Learning social etiquette: Human trajectory understanding in crowded scenes. In *Computer vision—ECCV 2016: 14th European conference, Amsterdam, The Netherlands, October 11–14, 2016, proceedings, Part VIII 14* (pp. 549–565). Springer.
- Search and rescue helicopter statistics: Year ending March 2021. <https://www.gov.uk/government/statistics/search-and-rescue-helicopter-annual-statistics-year-ending-march-2021/search-and-rescue-helicopter-statistics-year-ending-march-2021>
- Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*. <https://doi.org/10.1186/s40537-019-0197-0>
- Singh, B., & Davis, L. S. (2018). An analysis of scale invariance in object detection snip. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3578–3587).
- Singh, B., Najibi, M., & Davis, L. S. (2018). Sniper: Efficient multi-scale training. In *Advances in neural information processing systems* (Vol. 31).
- Stoddard, M. A., & Pelot, R. (2020). Historical maritime search and rescue incident data analysis. *Governance of Arctic Shipping: Rethinking Risk, Human Impacts and Regulation* (pp. 43–62).
- Sumbul, G., Charfuelan, M., Demir, B., & Markl, V. (2019). Bigearth-net: A large-scale benchmark archive for remote sensing image understanding. In *IGARSS 2019-2019 IEEE international geoscience and remote sensing symposium* (pp. 5901–5904). IEEE.
- Tan, M., Pang, R., & Le, Q. V. (2020). Efficientdet: Scalable and efficient object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*.
- U.S. Coast Guard Search and Rescue Statistics, Fiscal Year. <https://www.bts.gov/content/us-coast-guard-search-and-rescue-statistics-fiscal-year>
- Wang, Y., Bashir, S. M. A., Khan, M., Ullah, Q., Wang, R., Song, Y., Guo, Z., & Niu, Y. (2022). Remote sensing image super-resolution and object detection: Benchmark and state of the art. *Expert Systems with Applications*, 197, 116793.
- Wang, Y., Cai, Y., Liang, Y., Ding, H., Wang, C., Bhatia, S., & Hooi, B. (2021). Adaptive data augmentation on temporal graphs. *Advances in Neural Information Processing Systems*, 34, 1440–1452.
- Xia, G.-S., Bai, X., Ding, J., Zhu, Z., Belongie, S., Luo, J., Datcu, M., Pelillo, M., & Zhang, L. (2018). Dots: A large-scale dataset for object detection in aerial images. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3974–3983).
- Zhu, P., Wen, L., Du, D., Bian, X., Hu, Q., & Ling, H. (2020). Vision meets drones: Past, present and future 1(2), 8. *arXiv:2001.06303*
- Zhu, Q., Ma, K., Wang, Z., & Shi, P. (2023). Yolov7-csw for maritime target detection. *Frontiers in Neurorobotics* 17.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.