



## Article

# Automated Recognition of Submerged Body-like Objects in Sonar Images Using Convolutional Neural Networks

Yan Zun Nga, Zuhayr Rymansaib , Alfie Anthony Treloar and Alan Hunter \*

Faculty of Engineering and Design, University of Bath, Bath BA2 7AY, UK; yzn21@bath.edu (Y.Z.N.); z.rymansaib@bath.ac.uk (Z.R.); a.o.anthony.treloar@bath.ac.uk (A.A.T.)

\* Correspondence: a.j.hunter@bath.ac.uk

**Abstract:** The Police Robot for Inspection and Mapping of Underwater Evidence (PRIME) is an uncrewed surface vehicle (USV) currently being developed for underwater search and recovery teams to assist in crime scene investigation. The USV maps underwater scenes using sidescan sonar (SSS). Test exercises use a clothed mannequin lying on the seafloor as a target object to evaluate system performance. A robust, automated method for detecting human body-shaped objects is required to maximise operational functionality. The use of a convolutional neural network (CNN) for automatic target recognition (ATR) is proposed. SSS image data acquired from four different locations during previous missions were used to build a dataset consisting of two classes, i.e., a binary classification problem. The target object class consisted of 166  $196 \times 196$  pixel image snippets of the underwater mannequin, whereas the non-target class consisted of 13,054 examples. Due to the large class imbalance in the dataset, CNN models were trained with six different imbalance ratios. Two different pre-trained models (ResNet-50 and Xception) were compared, and trained via transfer learning. This paper presents results from the CNNs and details the training methods used. Larger datasets are shown to improve CNN performance despite class imbalance, achieving average F1 scores of 97% in image classification. Average F1 scores for target vs background classification with unseen data are only 47% but the end result is enhanced by combining multiple weak classification results in an ensemble average. The combined output, represented as a georeferenced heatmap, accurately indicates the target object location with a high detection confidence and one false positive of low confidence. The CNN approach shows improved object detection performance when compared to the currently used ATR method.

**Keywords:** underwater search; automation; robotics; sidescan sonar (SSS); automated target recognition (ATR); machine learning; convolutional neural networks (CNN)



**Citation:** Nga, Y.Z.; Rymansaib, Z.; Anthony Treloar, A.; Hunter, A. Automated Recognition of Submerged Body-like Objects in Sonar Images Using Convolutional Neural Networks. *Remote Sens.* **2024**, *16*, 4036. <https://doi.org/10.3390/rs16214036>

Academic Editors: Li Fang, Jian Yang, Yu Feng and Andrzej Stateczny

Received: 20 September 2024

Revised: 22 October 2024

Accepted: 26 October 2024

Published: 30 October 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Locating, identifying and recovering evidence from underwater environments is generally a manual, labour-intensive process reserved for specialist-trained divers [1]. Personnel involved in such operations have specific skill sets for proper and timely search procedures, as well as documentation and retrieval of any evidence discovered [2]. A common mission scenario is to locate and retrieve missing persons from inland waterways such as rivers or harbours [3,4]. Such environments can be dangerous to navigate due to poor underwater visibility, further complicated by potential hazards and obstructions littered in the area. Such time-consuming operations place considerable strain on increasingly over-stretched workforces with limited resources and personnel.

Sidescan sonar (SSS) is often employed in such scenarios to aid in underwater searches. However, it is still primarily a manual operation, requiring oversight from trained sonar operators [5–7]. Autonomous vessels have been developed for similar marine sensing scenarios [8,9], but automatic target recognition (ATR) of SSS data focuses primarily on applications such as mine hunting [10,11], seafloor characterisation [12–14], or general

man-made object and debris detection [15–17]. Autonomous detection of human-shaped targets has been demonstrated in multibeam sonar data [18,19] but in general, SSS is better suited for object detection, providing images of higher resolution and better quality [20,21].

The Police Robot for Inspection and Mapping of Underwater Evidence (PRIME) is an uncrewed surface vessel (USV) currently in development to assist in missing persons search and recovery scenarios. Detailed descriptions of the hardware design and software architecture are provided in the literature [22,23]. The USV is deployed to scan an area, producing SSS imagery of the floor. An ATR algorithm uses wavelet filtering and complexity mapping to identify human body-shaped objects, and produces simplified survey maps for dive teams in the form of a ‘heatmap’ for easy interpretation. The heatmap uses a two-colour scale, where benign regions are highlighted in blue, and areas highlighted in red indicate the potential presence of a body. This system has been shown to autonomously identify and locate a sunken mannequin, representing a human body in a missing persons scenario. However, the current image processing-based ATR approach incorrectly classifies water-land boundaries and changes in floor texture as regions of interest, resulting in several false positives as well as false negatives.

Convolutional neural networks (CNNs) are widely used for pattern recognition in tasks related to computer vision [24]. Recent work using CNNs has shown promising results in image classification [25,26] and object detection in sonar images [27,28]. Popular CNN architectures include Alexnet [29], VGGNets [30], ResNet [31], and Xception [32]. CNNs have been used extensively for object detection in sonar images [33–35]. Developing and training a complete CNN requires considerable time, computational resources, and training data. One approach to circumvent these requirements is to use transfer learning, where a pre-trained CNN is adapted for a new task by using the lower network layers to extract image features (e.g., edges, shapes, and textures) and retraining only the final output classification layer. This method has demonstrated excellent performance in applications ranging from Alzheimer’s detection in MRI images to flower species recognition from mobile phone photography [36–38], and has also been demonstrated to be suitable for object detection in SSS images [39–41].

In this paper, we describe a CNN approach for ATR. Two CNNs (ResNet-50 and Xception) were trained via transfer learning using data collected from previous missions. Most of the training data contain non-target imagery, resulting in an imbalanced class distribution. Basic data augmentation was thus performed, and the effect of varying the imbalance ratio on target versus background classification performance was examined. Results are further refined via ensemble averaging of classifications from multiple survey passes at different orientations. This minimises the effects of false negatives and results in improved target detection and localisation performance of the system.

## 2. Data Curation and Classifier Training

Traditional machine learning involves training models from scratch on specific datasets for a particular task. This typically requires significant computational resources and large amounts of good-quality training data, followed by testing and tuning to produce a model which can accurately classify new data. Transfer learning adapts existing models that have been trained on large datasets for general tasks and repurposes them for a specific task. The top-most classification layers are removed allowing the remaining pre-trained network to be used as a feature extractor. This reduces the requirement of using large datasets for training. New output layers are then added and trained on a smaller dataset to classify task-specific features, before fine-tuning of the whole network to refine overall performance.

Data collected from previous missions and testing were manually sorted and labelled to use for transfer learning with two CNN architectures, ResNet-50 [31] and Xception [32]. Two classes were specified as object and background, and datasets with varying imbalance ratios were prepared. Some data were withheld from training to be used later as unseen datasets after experimenting with different CNN architectures and imbalance ratios. After transfer learning, model performances were evaluated on full-size sonar images.

The unseen datasets were used to simulate new missions and evaluate object detection performance. Finally, CNN outputs were integrated into the PRIME autonomy chain to generate an intelligence map and compare against the existing object detection method.

### 2.1. Data Collection

The USV has been tested extensively during previous development stages at several locations between Bath and Bristol, UK. SSS data gathered from four locations were curated for training and testing the CNNs. This consists of sonar images generated using the higher frequency Starfish-990 SSS (Blueprint Subsea, UK) implemented on the USV platform. Table 1 lists details of each mission including location, number of survey runs, and number of sonar images collected in each survey. Underfall Yard is a boatyard serving Bristol Harbour with a water depth of approximately 5 m, see Figure 1. It is a moderately busy operational harbour with traffic from passing boats and sailing activities. The seafloor is soft and muddy with some clutter, rock formations, and texture. Bathampton and Dundas Aqueduct are two locations at different points along the Kennet and Avon Canal. Water is shallow, varying in depth from 1.2–1.5 m, with some clutter such as rocks, tyres, and other man-made objects on the muddy floor. Boats are moored at several locations along the length of the canal, and there is traffic from passing canal boats. Minerva Bath Rowing Club is a quiet area situated on the River Avon. The water is approximately 2.5 m deep, with a muddy floor and moderately complex underwater environment containing rocks, aquatic plants and fish. These locations are highlighted on a map in Figure 2.

**Table 1.** List of previous surveys indicating test location, date, number of survey runs, and number of sonar images collected. Surveys with fewer images are from incomplete runs. Data from highlighted lines were retained as unseen datasets for CNN evaluation.

No.	Location	Date	Run	No. of Images
1	Underfall Yard	02/10/2019	1	32
			2	24
		10/10/2019	1	70
			2	72
			3	68
		03/01/2020	1	20
			2	32
			3	36
		06/03/2020	1	140
			2	70
		14/07/2021	1	144
			2	42
2	Bathampton Canal	05/10/2017	1	130
3	Dundas Aqueduct	17/06/2021	1	100
4	Minerva Bath Rowing Club	09/07/2021	1	100
			2	68
			3	96



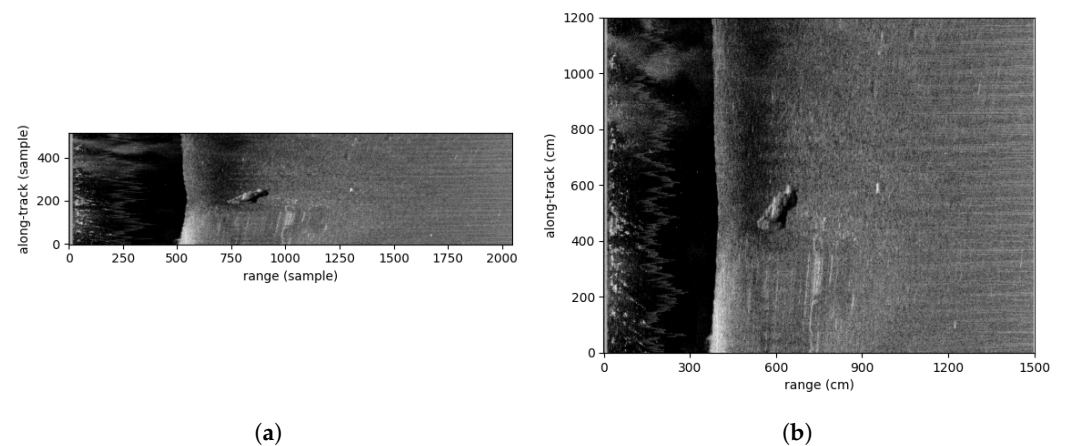
**Figure 1.** PRIME USV executing a survey at Underfall Yard, Bristol Harbour, UK.



**Figure 2.** Satellite map showing locations where experimental trials took place: (1) Bristol Harbour; (2) Bathampton Canal; (3) Dundas Aqueduct; (4) Minerva Bath Rowing Club.

## 2.2. Image Preparation

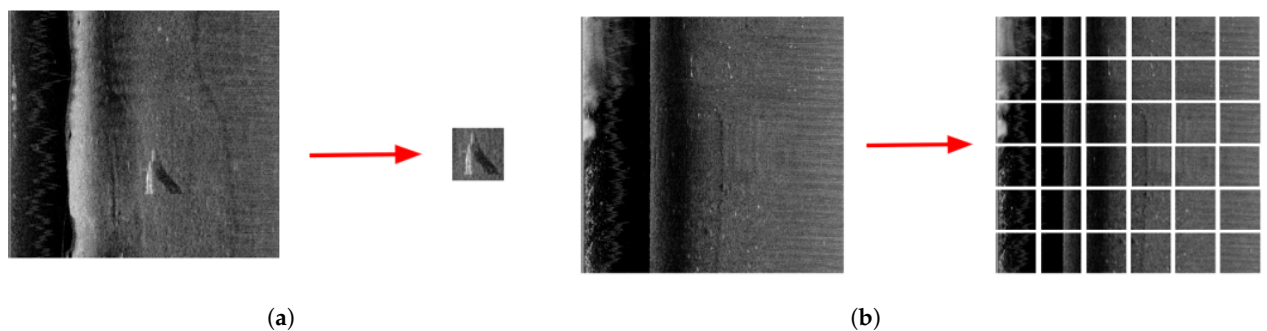
SSS data are collected at a fixed ping repetition frequency of 50 Hz as the USV travels along a path. Along-track sampling is thus determined by the USV velocity, which can vary due to changes in wind or water current. Acquired images, therefore, require rescaling to ensure object dimensions are consistent between images. Across-track sampling is assumed to be uniform, where each sonar ping covers a fixed range of 15 m using 2048 samples. Along-track samples are rescaled assuming a constant average velocity, with distance determined from the USV GPS data. Images were rescaled using bicubic interpolation (Figure 3), before manually sorting into two groups of object and background data.



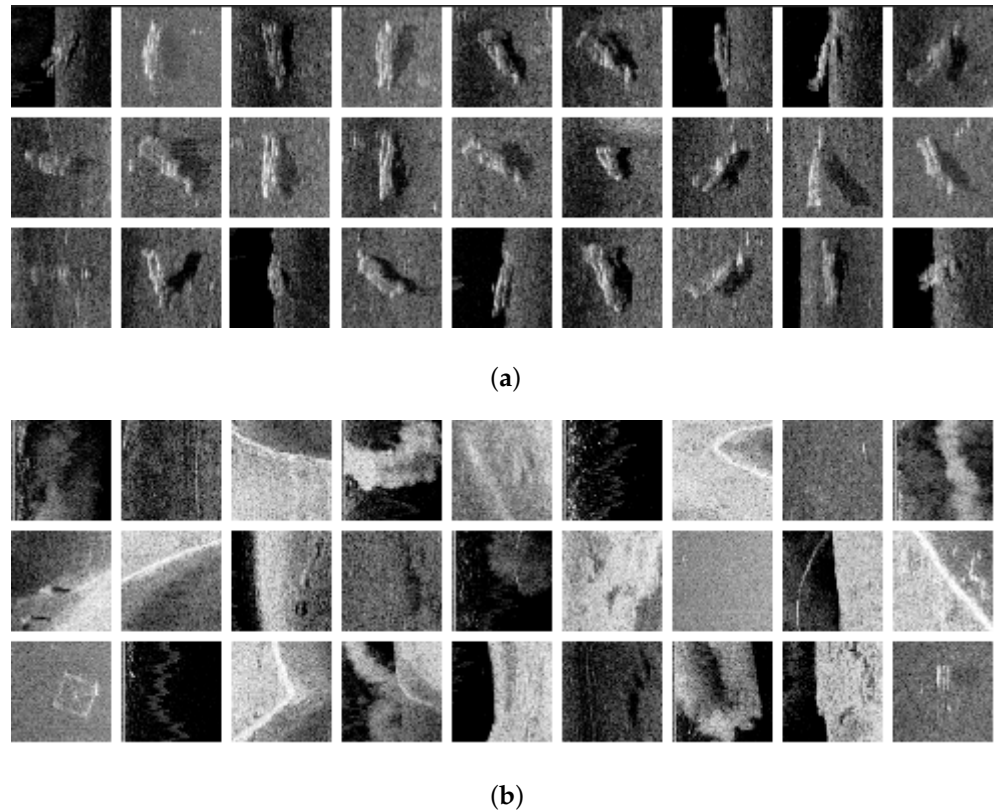
**Figure 3.** Example of a sonar image of the target object before (a) and after (b) rescaling using bicubic interpolation.

### 2.3. Data Labelling

Manual sorting of rescaled images resulted in 166 containing the target object and 680 without. Target images were cropped to  $196 \times 196$  pixels centred on the highlight and shadow features to generate samples for the object class with the remaining data discarded, see Figure 4a. A  $196 \times 196$  pixel sliding window with a 100 pixel step was used to generate samples for the background class from non-target images, see Figure 4b. This size was chosen to ensure the sliding window covered the entire sonar image. This resulted in a total of 166 samples for the object class, and 13,054 for the background class. To increase the amount of object class samples, basic data augmentation was performed by mirroring in the along-track direction, duplicating the amount of available data. Other common data augmentation techniques such as rotation cannot be used as the characteristic highlight-shadow feature would not be correctly reproduced. A selection of samples is shown in Figure 5, illustrating the wide variety of potential object configurations and background features.



**Figure 4.** Examples of generated training data: (a) Object class sample cropped from image containing the target object, remaining data are discarded. (b) Background class data generated from entire target-free image.



**Figure 5.** Gallery of example training data for object class (a) and background class (b).

#### 2.4. Transfer Learning

We perform transfer learning with two popular pre-trained networks, ResNet-50 [31] and Xception [32], both trained on the ImageNet dataset [42]. ResNet-50 is a 50 layer neural network which learns residual mappings between inputs and outputs, rather than directly learning the output. Shortcut connections propagate inputs from one layer to later layers, allowing the network to easily learn residual mappings without overfitting. The use of residual blocks in the architecture addresses the problem of vanishing gradients in deep neural networks. Xception is a 36 layer network that uses a more efficient version of convolution known as depthwise separable convolution. Image input channels are convolved separately and combined, which reduces the computation and number of parameters required whilst maintaining performance comparable to deeper networks.

Pre-trained networks were obtained using Keras an open-source library for neural networks written in Python [43]. The top layers were omitted and replaced with an average pooling layer followed by a 2 output fully connected layer with softmax activation. The new classifier head was then trained for 3 epochs with a batch size of 64 using the Adam optimiser with a learning rate of 0.0001 [44]. The entire model was then fine-tuned with the same parameters for 30 epochs or until early stopping, specified as a minimum change in validation loss of 0.001 over 3 epochs.

Several datasets were prepared to investigate the influence of class imbalance, see Table 2. These were created using all available samples from the object class and randomly selected samples from the background class. 20% of each dataset was assigned as testing data, and manually curated to ensure that samples from every mission were included. The remaining data were split 80:20 as training and validation data, respectively, with stratified random sampling, resulting in a total of 12 trained networks. This was repeated 10 times to evaluate the average model performance for each imbalance ratio, for both CNN architectures. Data from the missions highlighted in Table 1 were withheld from training to use as unseen data.

**Table 2.** List of training datasets with varying imbalance ratio.

No.	Object Examples	Background Examples	Imbalance Ratio
a	332	332	1:1
b		664	1:2
c		1328	1:4
d		2656	1:8
e		5312	1:16
f		10,624	1:32

### 2.5. Object Detection and Mapping

The trained networks were used to perform object detection on full-size sonar images from the testing data and from the unseen data. A total of 43 images from the testing data were used, all containing the target object. The unseen dataset consisted of 144 images, with 41 containing the target object. A  $196 \times 196$  pixel sliding window with a 30 pixel step was used to generate region proposals for classification. Non-maximum suppression with an overlap threshold of 0.2 was used to produce bounding boxes with the highest confidence rates around any detected objects [45]. Detections with confidence rates lower than 50% were ignored. Images with any placed bounding boxes were then converted to a heatmap with a blue-red colourscale where the bounded region is coloured red with the intensity determined by the classification confidence rate.

Heatmaps were then fed into the georeferencing and mapping stages as implemented in the PRIME processing chain, replacing the current wavelet filtering-based anomaly detection algorithm. The implementation is described in previous work [22]. Images are first filtered via wavelet transform to isolate features of similar scale to an average human body. The complexity of the filtered image is then quantified via the root-mean-square contrast metric to generate the anomaly map, where high complexity regions indicate an increased likelihood of the presence of the target object [46,47]. Finally, edge detection is performed to detect the seafloor and remove the water column region from sonar images and corresponding heatmaps.

## 3. Results and Discussion

Precision, recall, and F1 score [48] were used to evaluate the classification performances of both networks on the test datasets. The same metrics were used to evaluate object detection performance on images used to generate the test dataset and images from the unseen datasets. Figure 6 shows the distribution of F1 scores. Tables 3–5 show the averaged performances for classification of image snippets, object detection on images from the test data and object detection on unseen data, respectively.

**Table 3.** Averaged classification performance of Xception and ResNet-50 on test data samples (image snippets). Networks were trained and tested 10 times per imbalance ratio. Each dataset consisted of all available target object samples and a random subset of available background samples. The red-yellow-green colour scale ranges from 0–1 indicating poor to good performance.

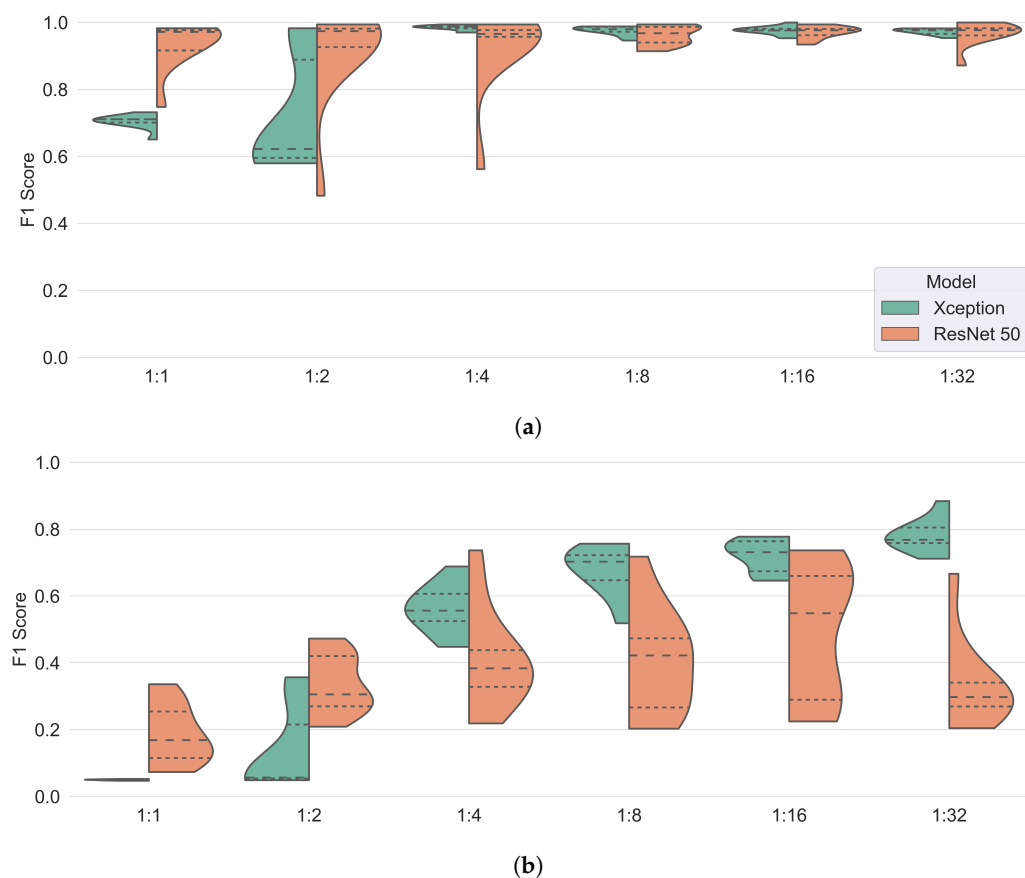
Model		Xception			ResNet-50		
No.	Imbalance Ratio	Precision	Recall	F1 Score	Precision	Recall	F1 Score
a	1:1	0.97	0.55	0.70	0.99	0.89	0.93
b	1:2	0.96	0.60	0.72	0.90	0.96	0.91
c	1:4	0.98	0.99	0.98	0.91	0.97	0.93
d	1:8	0.97	0.99	0.98	0.96	0.97	0.96
e	1:16	0.97	0.99	0.98	0.97	0.98	0.97
f	1:32	0.96	0.98	0.97	0.97	0.97	0.97

**Table 4.** Averaged object detection performance of Xception and ResNet-50 on full-size sonar images from the test dataset. The red-yellow-green colour scale ranges from 0–1 indicating poor to good performance.

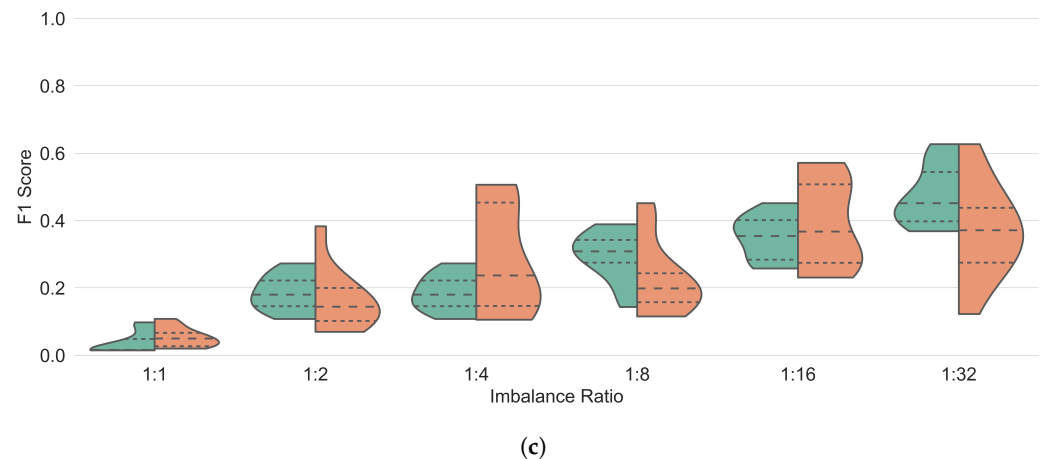
Model		Xception			ResNet-50		
No.	Imbalance Ratio	Precision	Recall	F1 Score	Precision	Recall	F1 Score
a	1:1	0.03	0.97	0.05	0.10	0.99	0.19
b	1:2	0.08	0.98	0.14	0.21	0.94	0.34
c	1:4	0.40	0.97	0.56	0.29	0.93	0.41
d	1:8	0.52	0.97	0.68	0.27	0.97	0.40
e	1:16	0.58	0.97	0.72	0.35	0.96	0.49
f	1:32	0.66	0.97	0.78	0.21	0.96	0.34

**Table 5.** Averaged object detection performance of Xception and ResNet-50 on full size sonar images from the unseen dataset. The red-yellow-green colour scale ranges from 0–1 indicating poor to good performance.

Model		Xception			ResNet-50		
No.	Imbalance Ratio	Precision	Recall	F1 Score	Precision	Recall	F1 Score
a	1:1	0.02	1.00	0.04	0.03	1.00	0.05
b	1:2	0.10	1.00	0.19	0.10	0.97	0.16
c	1:4	0.10	1.00	0.19	0.19	0.94	0.29
d	1:8	0.18	0.99	0.30	0.13	0.99	0.23
e	1:16	0.21	0.98	0.35	0.25	0.99	0.39
f	1:32	0.32	0.99	0.47	0.24	0.99	0.37



**Figure 6.** Cont.



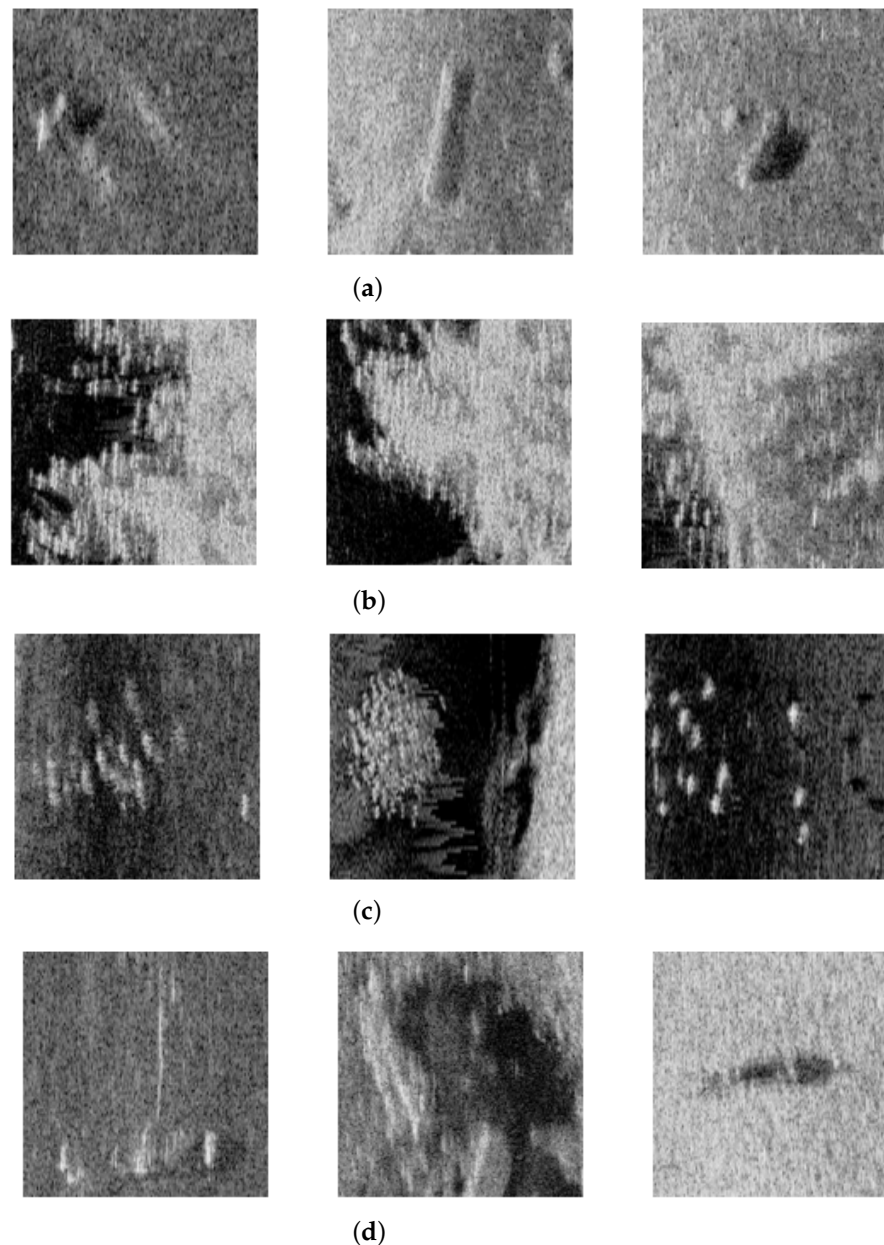
**Figure 6.** F1 scores of Xception and ResNet-50 for (a) classification on test data samples (image snippets), (b) object detection on full-size sonar images, (c) object detection on full-size sonar images from an unseen dataset, with increasing imbalance ratios. Lines show median and quartiles of distributions. Networks were re-trained 10 times each.

In most cases, Xception outperforms ResNet-50. We postulate that this is due to Xception being a smaller network and thus less prone to overfitting. Smaller networks have been shown to outperform larger CNNs in classifying sonar imagery [49]. Both networks tend to perform better overall when trained with higher imbalance ratios. This is likely due to a greater amount of data being used when training, allowing the network to more accurately model the background class. Figure 6a suggests the good performance of both models with the most imbalance ratios; however, due to the limited amount of available training data, k-fold cross-validation could not be properly performed and some data overlap between training, validation and test datasets could not be avoided during evaluation. This type of testing has been shown to be unreliable with small datasets [50]. A more representative assessment is shown in Figure 6c, where the best-performing models achieve F1 scores of 0.63 in object detection on unseen images. This moderate performance may also be due to the lack of sufficient training data. Additionally, the training and validation data used here have been collected from multiple trials and environments, yet the unseen data are from one environment only.

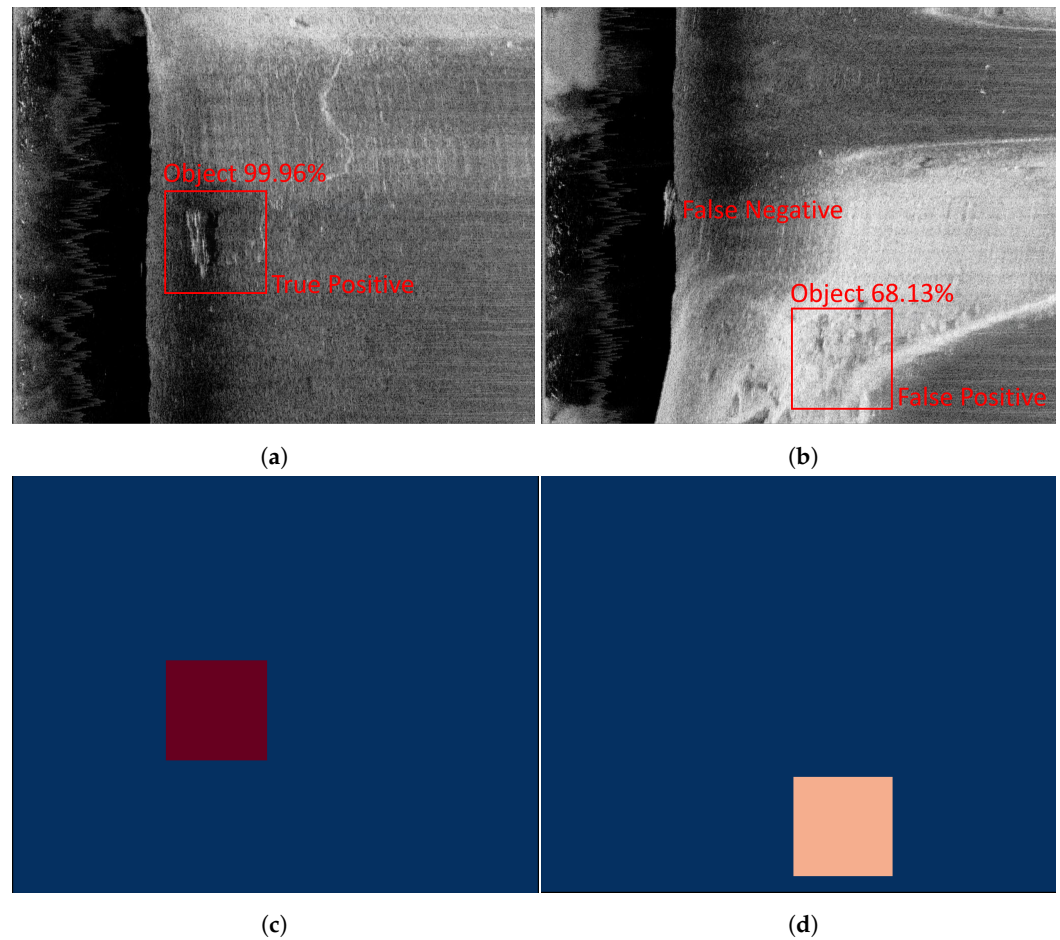
Some of the most common misclassifications are shown in Figure 7. Rocks (Figure 7a), plants (Figure 7b) and fish (Figure 7c) can be incorrectly identified as the object class, i.e., false positives. Plants and fish appear mainly in the water column, which is later removed in the final mapping stage, thereby reducing their impact as false positives. Figure 7d shows examples of false negatives, which are more likely to occur when the target object is near some clutter or is viewed from an unfavourable angle.

The best performing CNN from Table 5, i.e., Xception trained with an imbalance ratio of 1:32, was used to perform object detection on the unseen datasets. Detection examples and corresponding heatmaps are shown in Figure 8. Some detections are poor, incorrect, or missed altogether as seen in Figure 8b. With the average F1 score of this network on unseen data achieving only 0.47, the single look classification performance is poor. Each classification is view-dependent, with accuracy varying when viewing the scene from different orientations. Figure 9 illustrates how viewing the target object from an unfavourable orientation results in poor detection with the existing method, but nearby fish give a slightly stronger detection resulting in a false positive, see Figure 9b. The CNN-based approach results in a true positive detection for the target object but also incorrectly classifies the fish with high confidence, see Figure 9c. However, a good outcome is achieved by combining the results of multiple heatmaps, similar to an ensemble average of multiple weak classifications. Similar improvements to the classification of SSS images by using multiple looks of the same object have been demonstrated previously with dummy mines [51,52]. Figure 10b shows how the true target object position is clearly identified,

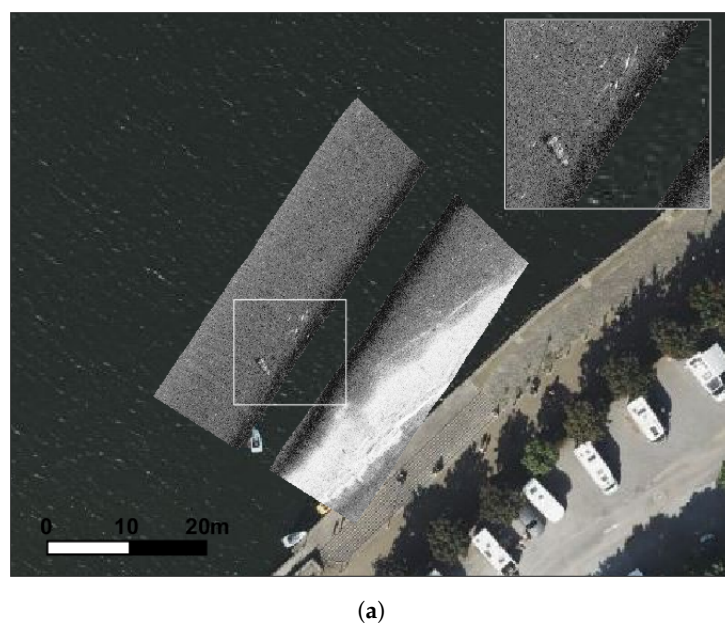
with the impact of false negatives arising from transient objects such as fish suppressed. A secondary region with overlapping false positives of low confidence is also present, due to rocks on the seafloor. This result is a significant improvement over the existing method shown in Figure 10a, which shows two additional regions of interest due to the algorithm incorrectly identifying changes in seafloor texture and the water-land boundary as anomalies. Finally, Figure 11 illustrates the good repeatability of this method with the additional unseen dataset, with the target location clearly identified when deployed in a different position. The location of the rocks is less pronounced in this result, this could be due to underwater changes which may have occurred in the 16-month period between gathering the two datasets.



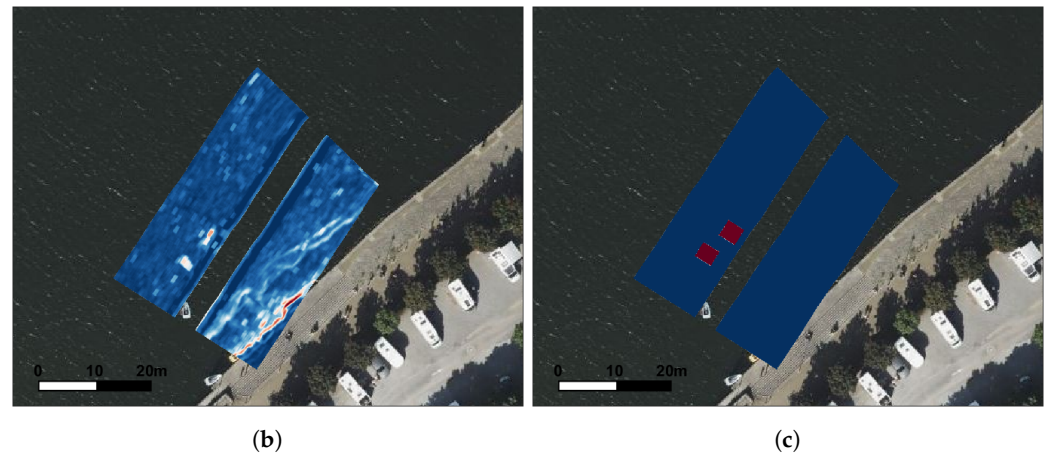
**Figure 7.** Examples of region proposals with the highest misclassification rates. (a) Rocks, (b) vegetation, and (c) fish were frequently misclassified as the target object. (d) Shows the target object misclassified as background.



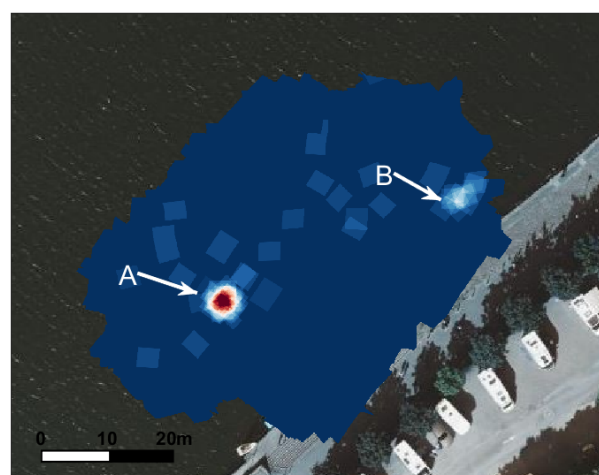
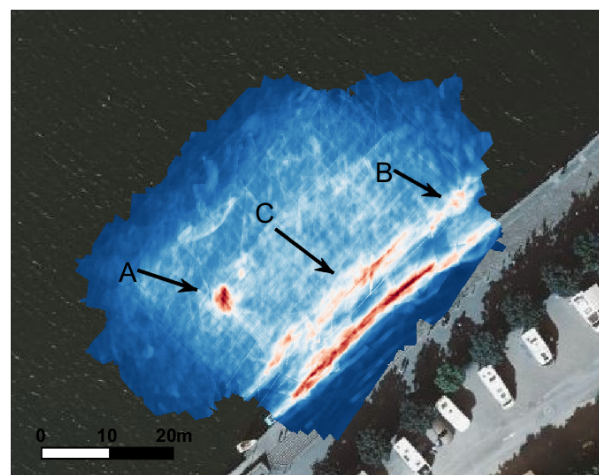
**Figure 8.** Examples of bounding box placement after object recognition. Left column (a,c) shows true positive detected with a high confidence rate, right column (b,d) shows false positive with moderate confidence rate and a false negative. Second row (c,d) shows corresponding heatmaps generated from the bounding box and confidence rate. Blue-white-red colour scale ranges from 0–100 and indicates confidence rate.



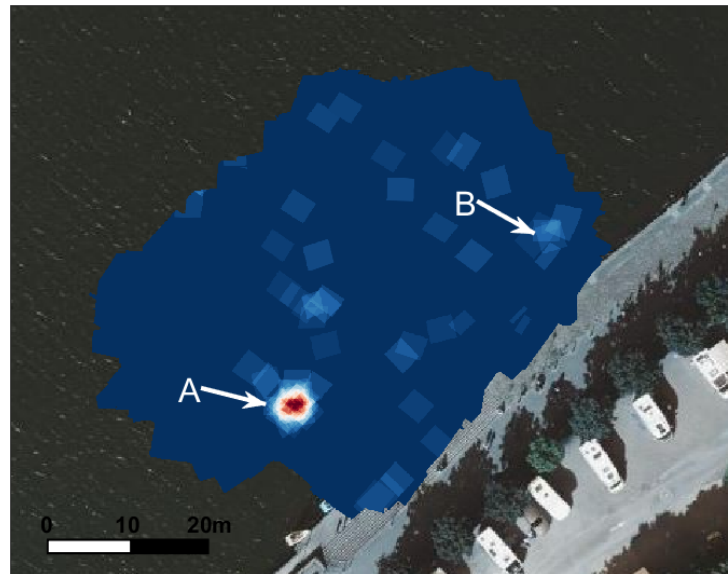
**Figure 9.** Cont.



**Figure 9.** Example of georeferenced (a) sonar image, with corresponding heatmaps generated using (b) existing ATR method and (c) CNN-based ATR method. Both methods incorrectly identify a school of fish as the target object. The blue-white-red colour scales in (b) and (c) range from 0.075–0.2 and 0–1, respectively.



**Figure 10.** Comparison of combined heatmaps produced using (a) existing image processing-based anomaly detection method and (b) CNN-based object detection method. The blue-white-red colour scales in (a) and (b) range from 0–0.015 0–0.3, respectively. Overlaid arrows indicate location of (A) target, (B) rocks and (C) change in floor texture. These results correspond to run 1 of the 2021 dataset highlighted in Table 1.



**Figure 11.** Combined heatmap produced using CNN-based object detection method on additional unseen dataset. Overlaid arrows indicate location of (A) target, (B) rocks shown in Figure 10. The blue-white-red colour scale ranges from 0–0.3. This result corresponds to run 1 of the 2020 dataset highlighted in Table 1.

#### 4. Conclusions

This paper presents two CNN architectures, ResNet-50 and Xception, trained via transfer learning for identifying body-like objects from sidescan sonar images collected from an autonomous USV in missing persons scenarios. A scarcity of available data motivated the investigation of training with varying imbalance ratios. In general, Xception outperformed ResNet-50 based on F1 classification scores. Networks trained with higher imbalance ratios achieved higher scores, suggesting that the negative class is more accurately modelled and that additional data for the object class are required. When evaluated for object detection with unseen datasets, both networks performed poorly on single-look classification, with average F1 scores for Xception and ResNet-50 achieving 0.47 and 0.37, respectively. However, overall performance is enhanced by integrating the CNN outputs into the USV mapping and visualisation processes, thereby combining the results on a georeferenced map as an ensemble average. The end result is a significant improvement over the current implementation with the target object clearly identified, allowing the system to produce intelligence of better quality for search and recovery teams. Additional improvements may be achieved by applying this methodology to other sonar image object detection methods, for example, Mondrian detection, Markov random field, or integral-image-based approaches [53–55]. However, improvements to the data-driven CNN method used here will likely arise simply from using additional training data as they are acquired.

The sliding window-based approach used to generate region proposals is simple but effective. More efficient methods such as the selective search algorithm could be explored and compared, as well as more modern CNN architectures such as YOLO or Efficient-Det [56–59]. Further improvements can be achieved by collecting more training data for better object and background classification, or for identifying commonly misclassified objects such as fish or rocks. Given the use case of search and rescue, detecting additional classes such as discarded weapons or hazards would be beneficial, but also requires more data for training, data collected from different environments, or data from additional sensor types. The sonar image data used for this work have been released [23], as well as the position data required for mapping, to replicate Figures 9–11. We encourage readers to experiment with different CNN architectures and object detection frameworks.

**Author Contributions:** Conceptualisation, A.H.; methodology, A.H. and Y.Z.N.; software, Y.Z.N.; validation, Z.R. and A.A.T.; formal analysis, Y.Z.N.; data curation, Y.Z.N. and Z.R.; writing—original draft preparation, Y.Z.N.; writing—review and editing, Y.Z.N., Z.R., A.A.T. and A.H.; supervision, A.H.; project administration, A.H.; funding acquisition, A.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by EPSRC grant EP/X030156/1.

**Data Availability Statement:** The data that support the findings of this study are openly available at <https://data.mendeley.com/preview/5w8k78brjw?a=96bdb8a3-ecc1-47dd-b977-24027e2278c0>, accessed on 20 October 2024 [23].

**Conflicts of Interest:** The authors declare no conflicts of interest

## Abbreviations

The following abbreviations are used in this manuscript:

ATR	automatic target recognition
CNN	convolutional neural network
PRIME	police robot for inspection and mapping of underwater evidence
SSS	sidescan sonar
USV	uncrewed surface vessel

## References

1. Becker, R.F.; Nordby, S.H.; Jon, J. *Underwater Forensic Investigation*; CRC Press: Boca Raton, FL, USA, 2013.
2. Erskine, K.L.; Armstrong, E.J. *Water-Related Death Investigation: Practical Methods and Forensic Applications*; CRC Press: Boca Raton, FL, USA, 2021.
3. Jahangir, R. Nicola Bulley: Lancashire Police Find Body in River Wyre. *BBC News*. Available online: <https://www.bbc.co.uk/news/uk-england-64697300> (accessed on 7 July 2023).
4. Brown, S. Police Find Body in Poole Harbour During Search for Missing 20-Year-Old. *Dorset Live*. Available online: <https://www.dorset.live/news/dorset-news/police-find-body-poole-harbour-8305294> (accessed on 7 July 2023).
5. Ruffell, A. Lacustrine flow (divers, side scan sonar, hydrogeology, water penetrating radar) used to understand the location of a drowned person. *J. Hydrol.* **2014**, *513*, 164–168. [CrossRef]
6. Schultz, J.J.; Healy, C.A.; Parker, K.; Lowers, B. Detecting submerged objects: The application of side scan sonar to forensic contexts. *Forensic Sci. Int.* **2013**, *231*, 306–316. [CrossRef] [PubMed]
7. Healy, C.A.; Schultz, J.J.; Parker, K.; Lowers, B. Detecting Submerged Bodies: Controlled Research Using Side-Scan Sonar to Detect Submerged Proxy Cadavers. *J. Forensic Sci.* **2015**, *60*, 743–752. [CrossRef] [PubMed]
8. Moulton, J.; Karapetyan, N.; Bukhsbaum, S.; McKinney, C.; Malebary, S.; Sophocleous, G.; Li, A.Q.; Rekleitis, I. An autonomous surface vehicle for long term operations. In Proceedings of the OCEANS 2018 MTS/IEEE Charleston, Charleston, SC, USA, 22–25 October 2018; pp. 1–10.
9. Smith, T.; Mukhopadhyay, S.; Murphy, R.R.; Manzini, T.; Rodriguez, I. Path Coverage Optimization for USV with Side Scan Sonar for Victim Recovery. In Proceedings of the 2022 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR), Sevilla, Spain, 8–10 November 2022; pp. 160–165. [CrossRef]
10. Chapple, P.B. Unsupervised detection of mine-like objects in seabed imagery from autonomous underwater vehicles. In Proceedings of the OCEANS 2009, Biloxi, MS, USA, 26–29 October 2009; pp. 1–6. [CrossRef]
11. Gebhardt, D.; Parikh, K.; Dzieciuch, I.; Walton, M.; Vo Hoang, N.A. Hunting for naval mines with deep neural networks. In Proceedings of the OCEANS 2017—Anchorage, Anchorage, AK, USA, 18–21 September 2017; pp. 1–5.
12. Hamilton, L. Towards autonomous characterisation of side scan sonar imagery for seabed type by unmanned underwater vehicles. In Proceedings of the Proceedings of ACOUSTICS, Perth, Australia, 19–22 November 2017; pp. 1–10.
13. Nian, R.; Zang, L.; Geng, X.; Yu, F.; Ren, S.; He, B.; Li, X. Towards characterizing and developing formation and migration cues in seafloor sand waves on topology, morphology, evolution from high-resolution mapping via side-scan sonar in autonomous underwater vehicles. *Sensors* **2021**, *21*, 3283. [CrossRef]
14. Williams, D.P. Fast unsupervised seafloor characterization in sonar imagery using lacunarity. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 6022–6034. [CrossRef]
15. Fakiris, E.; Papatheodorou, G.; Geraga, M.; Ferentinos, G. An Automatic Target Detection Algorithm for Swath Sonar Backscatter Imagery, Using Image Texture and Independent Component Analysis. *Remote Sens.* **2016**, *8*, 373. [CrossRef]
16. Rhinelander, J. Feature extraction and target classification of side-scan sonar images. In Proceedings of the 2016 IEEE Symposium Series on Computational Intelligence (SSCI), Athens, Greece, 6–9 December 2016; pp. 1–6. [CrossRef]

17. Merrifield, S.T.; Celona, S.; McCarthy, R.A.; Pietruszka, A.; Batchelor, H.; Hess, R.; Nager, A.; Young, R.; Sadorf, K.; Levin, L.A.; et al. Wide-Area Debris Field and Seabed Characterization of a Deep Ocean Dump Site Surveyed by Autonomous Underwater Vehicles. *Environ. Sci. Technol.* **2023**, *57*, 18162–18171. [\[CrossRef\]](#)
18. Nguyen, H.T.; Lee, E.H.; Lee, S. Study on the classification performance of underwater sonar image classification based on convolutional neural networks for detecting a submerged human body. *Sensors* **2019**, *20*, 94. [\[CrossRef\]](#)[\[PubMed\]](#)
19. Lee, S.; Park, B.; Kim, A. A Deep Learning based Submerged Body Classification Using Underwater Imaging Sonar. In Proceedings of the 2019 16th International Conference on Ubiquitous Robots (UR), Jeju, Republic of Korea, 24–27 June 2019; pp. 106–112. [\[CrossRef\]](#)
20. Bates, C.R.; Lawrence, M.; Dean, M.; Robertson, P. Geophysical Methods for Wreck-Site Monitoring: The Rapid Archaeological Site Surveying and Evaluation (RASSE) programme. *Int. J. Naut. Archaeol.* **2011**, *40*, 404–416. [\[CrossRef\]](#)
21. Smith, C.J.; Rumohr, H. Imaging Techniques. In *Methods for the Study of Marine Benthos*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2013; Chapter 3, pp. 97–124. [\[CrossRef\]](#)
22. Rymansaib, Z.; Thomas, B.; Treloar, A.A.; Metcalfe, B.; Wilson, P.; Hunter, A. A prototype autonomous robot for underwater crime scene investigation and emergency response. *J. Field Robot.* **2023**, *40*, 983–1002. [\[CrossRef\]](#)
23. Rymansaib, Z.; Nga, Y.; Treloar, A.A.; Hunter, A. Sidescan sonar images for training automated recognition of submerged body-like objects. *Univ. Bath Res. Data Arch.* **2024**. [\[CrossRef\]](#)
24. Rawat, W.; Wang, Z. Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. *Neural Comput.* **2017**, *29*, 2352–2449. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Wang, X.; Jiao, J.; Yin, J.; Zhao, W.; Han, X.; Sun, B. Underwater sonar image classification using adaptive weights convolutional neural network. *Appl. Acoust.* **2019**, *146*, 145–154. [\[CrossRef\]](#)
26. Li, C.; Ye, X.; Cao, D.; Hou, J.; Yang, H. Zero shot objects classification method of side scan sonar image based on synthesis of pseudo samples. *Appl. Acoust.* **2021**, *173*, 107691. [\[CrossRef\]](#)
27. Jiang, L.; Cai, T.; Ma, Q.; Xu, F.; Wang, S. Active Object Detection in Sonar Images. *IEEE Access* **2020**, *8*, 102540–102553. [\[CrossRef\]](#)
28. Karimanzira, D.; Renkewitz, H.; Shea, D.; Albiez, J. Object Detection in Sonar Images. *Electronics* **2020**, *9*, 1180. [\[CrossRef\]](#)
29. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2012**, *60*, 84–90. [\[CrossRef\]](#)
30. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2015**, arXiv:1409.1556.
31. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
32. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1800–1807.
33. Thanh Le, H.; Phung, S.L.; Chapple, P.B.; Bouzerdoun, A.; Ritz, C.H.; Tran, L.C. Deep Gabor Neural Network for Automatic Detection of Mine-like Objects in Sonar Imagery. *IEEE Access* **2020**, *8*, 94126–94139. [\[CrossRef\]](#)
34. Zhang, F.; Zhang, W.; Cheng, C.; Hou, X.; Cao, C. Detection of Small Objects in Side-Scan Sonar Images Using an Enhanced YOLOv7-Based Approach. *J. Mar. Sci. Eng.* **2023**, *11*, 2155. [\[CrossRef\]](#)
35. Ge, L.; Singh, P.; Sadhu, A. Advanced deep learning framework for underwater object detection with multibeam forward-looking sonar. *Struct. Health Monit.* **2024**, 14759217241235637. [\[CrossRef\]](#)
36. Afzal, S.; Maqsood, M.; Nazir, F.; Khan, U.; Aadil, F.; Awan, K.M.; Mehmood, I.; Song, O.Y. A data augmentation-based framework to handle class imbalance problem for Alzheimer’s stage detection. *IEEE Access* **2019**, *7*, 115528–115539. [\[CrossRef\]](#)
37. Phong, T.D.; Duong, H.N.; Nguyen, H.T.; Trong, N.T.; Nguyen, V.H.; Van Hoa, T.; Snasel, V. Brain Hemorrhage Diagnosis by Using Deep Learning. In Proceedings of the 2017 International Conference on Machine Learning and Soft Computing, Ho Chi Minh City, Vietnam, 13–16 January 2017; pp. 34–39. [\[CrossRef\]](#)
38. Gogul, I.; Kumar, V.S. Flower species recognition system using convolution neural networks and transfer learning. In Proceedings of the 2017 Fourth International Conference on Signal Processing, Communication and Networking (ICSCN), Chennai, India, 16–18 March 2017; pp. 1–6. [\[CrossRef\]](#)
39. Ye, X.; Li, C.; Zhang, S.; Yang, P.; Li, X. Research on Side-scan Sonar Image Target Classification Method Based on Transfer Learning. In Proceedings of the OCEANS 2018 MTS/IEEE Charleston, Charleston, SC, USA, 22–25 October 2018; pp. 1–6. [\[CrossRef\]](#)
40. Ge, Q.; Ruan, F.; Qiao, B.; Zhang, Q.; Zuo, X.; Dang, L. Side-Scan Sonar Image Classification Based on Style Transfer and Pre-Trained Convolutional Neural Networks. *Electronics* **2021**, *10*, 1823. [\[CrossRef\]](#)
41. Du, X.; Sun, Y.; Song, Y.; Sun, H.; Yang, L. A Comparative Study of Different CNN Models and Transfer Learning Effect for Underwater Object Classification in Side-Scan Sonar Images. *Remote Sens.* **2023**, *15*, 593. [\[CrossRef\]](#)
42. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
43. Chollet, F. Keras. 2015. Available online: <https://keras.io> (accessed on 1 February 2021).
44. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
45. Hosang, J.; Benenson, R.; Schiele, B. Learning non-maximum suppression. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4507–4515.

46. Geilhufe, M.; Midtgaard, Ø. Quantifying the complexity in sonar images for MCM performance estimation. In Proceedings of the 2nd International Conference and Exhibition on Underwater Acoustics, Rhodes, Greece, 22–27 June 2014; pp. 1041–1048.
47. Peli, E. Contrast in complex images. *JOSA A* **1990**, *7*, 2032–2040. [[CrossRef](#)]
48. Sokolova, M.; Japkowicz, N.; Szpakowicz, S. Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation. In *AI 2006: Advances in Artificial Intelligence*; Sattar, A., Kang, B.H., Eds.; Springer: Berlin/Heidelberg, Germany, 2006; pp. 1015–1021.
49. Williams, D.P. On the Use of Tiny Convolutional Neural Networks for Human-Expert-Level Classification Performance in Sonar Imagery. *IEEE J. Ocean. Eng.* **2021**, *46*, 236–260. [[CrossRef](#)]
50. Isaksson, A.; Wallman, M.; Göransson, H.; Gustafsson, M.G. Cross-Validation and Bootstrapping Are Unreliable in Small Sample Classification. *Pattern Recogn. Lett.* **2008**, *29*, 1960–1965. [[CrossRef](#)]
51. Fawcett, J.; Myers, V.; Hopkin, D.; Crawford, A.; Couillard, M.; Zerr, B. Multiaspect Classification of Sidescan Sonar Images: Four Different Approaches to Fusing Single-Aspect Information. *IEEE J. Ocean. Eng.* **2010**, *35*, 863–876. [[CrossRef](#)]
52. Zerr, B.; Stage, B.; Guerrero, A. *Automatic Target Classification Using Multiple Sidescan Sonar Images of Different Orientations*; Technical Report; NATO, SACLANT Undersea Research Centre: La Spezia, Italy, 1997.
53. Williams, D.P. The Mondrian Detection Algorithm for Sonar Imagery. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 1091–1102. [[CrossRef](#)]
54. Reed, S.; Petillot, Y.; Bell, J. An automatic approach to the detection and extraction of mine features in sidescan sonar. *IEEE J. Ocean. Eng.* **2003**, *28*, 90–105. [[CrossRef](#)]
55. Williams, D.P. Fast Target Detection in Synthetic Aperture Sonar Imagery: A New Algorithm and Large-Scale Performance Analysis. *IEEE J. Ocean. Eng.* **2015**, *40*, 71–92. [[CrossRef](#)]
56. Uijlings, J.; van de Sande, K.; Gevers, T.; Smeulders, A. Selective Search for Object Recognition. *Int. J. Comput. Vis.* **2013**, *104*, 154–171. [[CrossRef](#)]
57. Redmon, J.; Divvala, S.K.; Girshick, R.B.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. *arXiv* **2015**, arXiv:1506.02640.
58. Wang, A.; Chen, H.; Liu, L.; Chen, K.; Lin, Z.; Han, J.; Ding, G. YOLOv10: Real-Time End-to-End Object Detection. *arXiv* **2024**, arXiv:2405.14458.
59. Tan, M.; Pang, R.; Le, Q.V. EfficientDet: Scalable and Efficient Object Detection. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 10778–10787. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.