Technical Note

Robust and Fair Undersea Target Detection with Automated Underwater Vehicles for Biodiversity Data Collection

Ranjith Dinakaranr [1], Li Zhang [2], Chang-Tsun Li [3], Ahmed Bouridane [4] and Richard Jiang [5,*]

1   Department of Computer and Information Sciences, Northumbria University, Newcastle NE1 8ST, UK; ranjith.dinakaran@northumbria.ac.uk
2   Department of Computer Science, Royal Holloway University of London, Surrey TW20 0EX, UK; li.zhang@rhul.ac.uk
3   School of Info Technology, Deakin University, Deakin, VIC 3125, Australia; changtsun.li@deakin.edu.au
4   Cybersecurity and Data Analytics Research Center, University of Sharjah, Sharjah P.O. Box 27272, United Arab Emirates; abouridane@sharjah.ac.ae
5   School of Computing and Communication, Lancaster University, Lancaster LA1 4YW, UK
*   Correspondence: r.jiang2@lancaster.ac.uk

Abstract: Undersea/subsea data collection via automated underwater vehicles (AUVs) plays an important role for marine biodiversity research, while it is often much more challenging than the data collection above ground via satellites or AUVs. To enable the automated undersea/subsea data collection system, the AUVs are expected to be able to automatically track the objects of interest through what they can "see" from their mounted underwater cameras, where videos or images could be drastically blurred and degraded in underwater lighting conditions. To solve this challenge, in this work, we propose a cascaded framework by combining a DCGAN (deep convolutional generative adversarial network) with an object detector, i.e., single-shot detector (SSD), named DCGAN+SSD, for the detection of various underwater targets from the mounted camera of an automated underwater vehicle. In our framework, our assumption is that DCGAN can be leveraged to alleviate the impact of underwater conditions and provide the object detector with a better performance for automated AUVs. To optimize the hyperparameters of our models, we applied a particle swarm optimization (PSO)-based strategy to improve the performance of our proposed model. In our experiments, we successfully verified our assumption that the DCGAN+SSD architecture can help improve the object detection toward the undersea conditions and achieve apparently better detection rates over the original SSD detector. Further experiments showed that the PSO-based optimization of our models could further improve the model in object detection toward a more robust and fair performance, making our work a promising solution for tackling the challenges in AUVs.

Keywords: automated underwater vehicles; biodiversity; object detection; deep neural networks; particle swarm optimization

1. Introduction

Undersea/subsea exploration is an important task in many industries, ranging from fishery to oil drilling, where undersea/subsea biodiversity [1] remains a concern due to the impact of these industry activities. The supply of safe and sustainable seafoods [2] is always an important factor for policymakers in many countries because food supply heavily relies on fishery industry. In particular, the recent environmental challenges such as global climate changes [3], arctic iceberg melting [4], microplastic pollution [5], and algae pollution [6] have become a surging threat to human life on the planet. In a long-term view, the study on undersea/subsea biodiversity not only has important economic and environmental value in single countries, but also a wide impact on the global environment and climate change.

However, unlike the investigation of land-cover-based biodiversity, subsea/undersea biodiversity [7] is not easy to directly study due to the difficulty in collecting data under the surface of seawater. Automated underwater vehicles (AUVs) [8–10] are an effective means for undersea and subsea investigation. Although it is technically feasible to use AUVs for underwater investigation, it is challenging for AUVs to maneuver automatically in deep sea due to the difficulties in signal communication [1], sensory data collection, and automation complexities. In particular, due to the murkiness in underwater conditions, the cameras mounted on AUVs may suffer from challenges such as lower quality of images/videos. Such challenges will not only degrade the performance of computer vision-based object detectors, but also lead to a biased detection by missing some specific targets in deep and dark waters.

Underwater object detection is an indispensable task in many automated AUV tasks such as the monitoring of the Aberdeen oil field, as shown in Figure 1a, where they can be used to fix various system failure issues. They can also be used for investigations of different species of fish underwater with blurred vision, as shown in Figure 1b. Underwater object detection is a complicated task [7], facing challenges due to the low light environment, water diffraction, water wave impact, etc. Consequently, these challenges lead to low-quality images or videos captured by the AUV-mounted cameras, where object detection becomes a tough task in such environments [8]. The work carried out in this paper is aimed at helping to alleviate underwater object detection using AUV-mounted cameras, while achieving robust and fair performance in object detection.

(**a**)	(**b**)
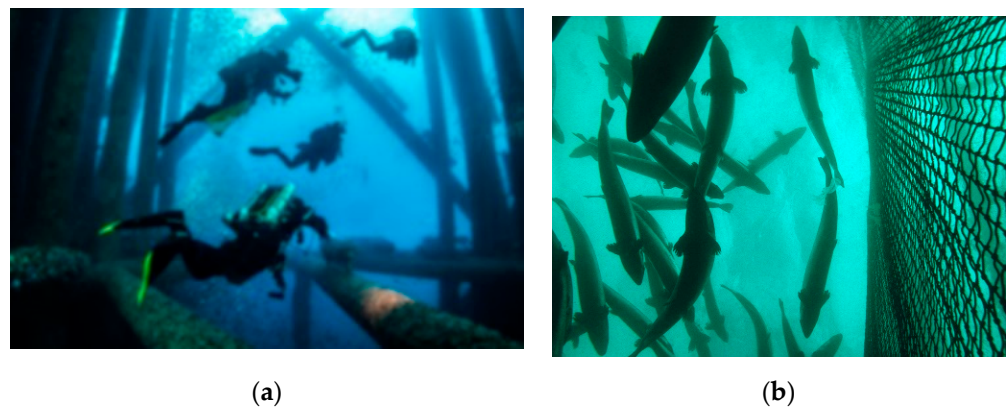
Figure 1. Undersea/subsea exploration: (**a**) Aberdeen oil field in the North Sea; (**b**) fish in Scotland sea farms.

In our recent work [11], it was evidenced that image quality can critically impact on object detection. In [12], we exploited a DCGAN (deep convolutional generative adversarial network) [13] for image enhancement, as shown in Figure 2, and we found that such an enhancement scheme could help alleviate the serious degradation of object detection caused by low-quality images. In this paper, we aimed to leverage our initial discovery and apply such a DCGAN-based framework [12] for undersea/subsea object detection using a single-shot detector (SSD) [14]. This combinational framework DCGAN+SSD can help object detection in underwater environments by engaging DCGAN for image conversion, after which the object detector is applied to the converted images. The idea is that such a combined structure would achieve a better accuracy on object detection in undersea videos, and the aim of this study was to validate this suggestion.
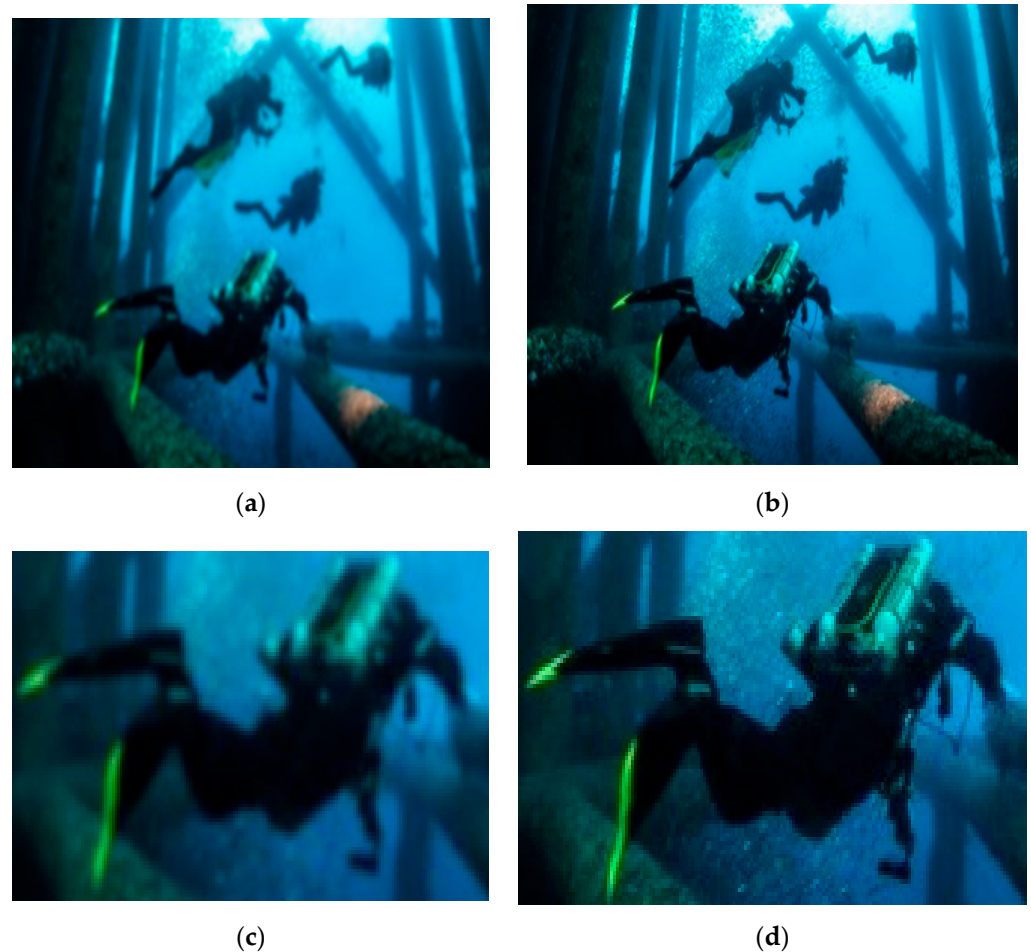
Figure 2. Image conversion toward the challenges in underwater conditions: (**a**) the original image; (**b**) the converted image by DCGAN; (**c**) an object in the original image; (**d**) an object in the converted image.

Here, we chose the SSD as our object detector because it is more robust in comparison with YOLO detectors [15,16]. YOLO detectors, as denoted by their name, "You Only Look Once", are more aimed at quick detection, whereas we were more concerned with accuracy for our AUV applications. Our goal is to tackle the challenges of hostile underwater conditions and enable the computer vision detector to identify various objects including humans, fishes, plants, and marine vertebrates. For biodiversity studies in particular, various object types may be targeted [7], and a robust and fair computer vision framework can greatly facilitate AUVs to fulfill a wide range of tasks, especially when AUVs require a smart self-piloting control for challenging civic, military, or scientific tasks.

In this paper, to adjust the DCGAN+SSD detector [12] for undersea/subsea environments, we propose a particle swarm optimization (PSO)-based optimization to tune the hyperparameters in the DCGAN+SSD framework. In our work, we used multiple datasets including the CIFAR-100, CADDY, and Roboflow fish datasets for PSO-based hyperparameter tuning. The tests were then carried out on various videos taken in the wild from undersea/subsea setups, implying a very challenging cross-dataset validation.

Our contribution mainly resides in two aspects. First, via our experiments, we successfully validated our assumption and demonstrated that the proposed DCGAN+SSD could enable object detection at better detection rates in undersea/subsea videos under hostile undersea conditions. Secondly, in our further experiments, we successfully proved that our PSO+DCGAN+SSD could work better than the initial DCGAN+SSD model, achieving much more robust and fair object detection toward undersea/subsea challenges.

In the remainder of the paper, Section 2 provides a literature review, Section 3 presents our DCGAN+SSD framework, and Section 4 describes our proposed PSO-based optimization method for hyperparameter tuning. Section 5 gives our experiments and validation. Section 6 concludes the paper.

2. Background and Related Work

Within the topics of video analysis and image understanding, object detection is a common task that has attracted great attention in recent years [14]. The problem is often modeled in many traditional and modern object detection methods as examining whether a targeted object of an expected class is present in a rectangle of regions in an image or a video frame. Despite the problems faced by object detectors and computer vision, object detection is a mandatory tool for many applications [17–22]. It is used in satellite image analysis [17], AI-automated medical diagnosis [18–20], robotic vision [21], behavioral analysis [22], 3D sensing [23], etc. In recent years, convolutional neural networks (CNNs) [24] have become the most popular technology in the field of computer vision, especially for object detection [15]. CNNs have been successfully exploited in many application fields such as self-driving cars [25], pedestrian detection [26] in smart city surveillance, and underwater object detection [10], as discussed in this work.

Object detection techniques were researched widely in recent years, and many methods have been recommended to effectively solve the object detection challenge [15]. Typical techniques dealing with object detectors include faster RCNN [15], SSD [12], and YOLO [14]. In the case of object detection, images are analyzed under a group of candidate regions [15] that could contain objects in an image. In [27], stacked convolution networks were exploited for object detection. In [28], object detection was modeled by detecting various components. Szegedy et al. [24] used DCGAN as a colorization technique for monochromatic problems. In this work, we aimed to exploit DCGAN for the challenges faced in underwater object detection.

2.1. Challenges in Underwater Object Detection: Robustness and Fairness

Underwater object detection using AUVs is a challenging task due to the hostile condition in undersea/subsea waters, as shown in Figure 1a,b. Oil drills, fishery monitoring, and biodiversity studies are examples [1–7] that urgently need a robust solution. As undersea/subsea tasks [8–10] are often expensive and dangerous for human divers, underwater object detection using AUVs is becoming more and more popular for various applications.

For example, in [1], object detection was applied to detect marine life with different colors and shapes for a biodiversity study, whereas, in [5], the focal loss suffered by underwater objects in front of cameras was studied. In this research, targeting the challenges in undersea conditions, we exploited DCGAN to improve the camera outputs, enabling a better fit for self-adjusted learning [10] in undersea object detection.

It is worth mentioning that we need to consider not only the robustness of the object detection, but also the fairness of the detector toward all targeted objects [29]. For example, in a biodiversity investigation [6,7], a biased detector may achieve very high detection rates on common objects and miss many rare species due to either data scarcity or, more simply, low-quality images in undersea conditions. As is often the case, biodiversity studies are more concerned about those rare species; hence, missing them could lead to incorrect inferences in biodiversity modeling.

2.2. Image Conversion via DCGAN

When considering the issues of image quality in object detection, DCGAN [12,13] can be a workable model to alleviate the challenges facing underwater object detection using AUVs. The application of DCGAN allows better learning of features from image pixels, which can make it easier for a detector to find objects in low-quality images or videos. DCGAN is used to enhance the data and features from objects to make them suitable for detection [30]. Another main reason for using DCGAN is its conversion from the

subspaces of image data, as mentioned by [31], with a potential dimensionality projection for a better understanding, particularly with respect to hyperspectral images from infrared or night-vision imaging sensors.

Experimental results have shown that the DCGAN framework is able to synthesize realistic images with diversity [32]. It can be used to add colorization in places where full color is lacking, whereby the DCGAN model can learn to colorize the lacking areas [33]. By enhancing the colors, it becomes possible to better identify the main features within an image, as lacking colors often lead to missing features, creating the possibility of misdetection. As an example, DCGAN was used to enhance the data in images of tomato leaf disease [34], with data augmentation being used to recreate the actual data instead of relying on synthetic data.

With the help of DCGAN, it has been suggested that a detector may work better over converted features [12]. Following our initial work in [12], further research was carried out and discussed in [15,26,35], and this DCGAN + detector framework was validated in many similar systems such as advanced driver assistance systems [25]. Similar work was reported by [33], where a novel small ship detection method was proposed using GAN and YOLOv2. A modified Wassertein generative adversarial network (WGAN) was used in [34] with a gradient penalty for plant disease detection. To solve the issues in synthetic images for traffic sign detection [32], complicated images were created using DCGAN, with a small amount of data stored as a solution for a synthetic image. A low-light environment is often encountered in our day-to-day life, and it is evident, in many cases, that low light can have a negative impact on object detection [35]. DCGAN can help alleviate the challenges in night object detection [28]. Chen et al. [33] used GAN for a semi-supervised learning method to deal with both labeled and unlabeled data via semi-supervised learning to extract useful information from labeled and unlabeled data to achieve a reasonable classifier, which can be extrapolated to an object detection task.

2.3. Balancing Dataset

In addition to issues of image quality [11,12], another reason for the choice of DCGAN in our work was to handle the unbalanced dataset [25,28,32,33]. While deep learning is well known for its dependence on a large amount of data, data collection can be heavily unbalanced, suffering from data scarcity in some key cases. In object detection for AUVs, the data from an undersea environment may not be sufficient compared to normal conditions above ground. To address this problem, in our research, we utilized DCGAN [13] to convert the undersea images, with the aim of making the images more suitable for the object detector, thus providing an alternative solution to the possible issues of imbalanced data [36]. With such an expectation, we examined, via our experiments, the assumption that DCGAN could alleviate the issues of data scarcity that may undermine the robustness and fairness in object detection.

3. Our DCGAN+SSD Framework

Object detection is considered a major challenge in computer vision despite some success in recent years thanks to recent advances in deep learning technologies [15]. Among different deep learning architectures, DCGAN is one of most interesting [12,28,30–36]. The standard architecture of a DCGAN often consists of two parts: a generator which can produce style-enhanced images from low-quality inputs [12], and a discriminator which aims to verify if the quality of converted images matches with the criteria according to the input [34].

To utilize DCGAN in object detection, we cascaded it with an object detector, as shown in Figure 3. Here, we used the SSD [14] as our object detector. Among other candidates, faster RCNN [15] is a bit older, whereas YOLO [16] focuses on achieving a low computing time. Our assumption is that such a combination could alleviate the challenges in undersea object detection, which are likely caused by the low image quality and the scarcity of undersea data samples.
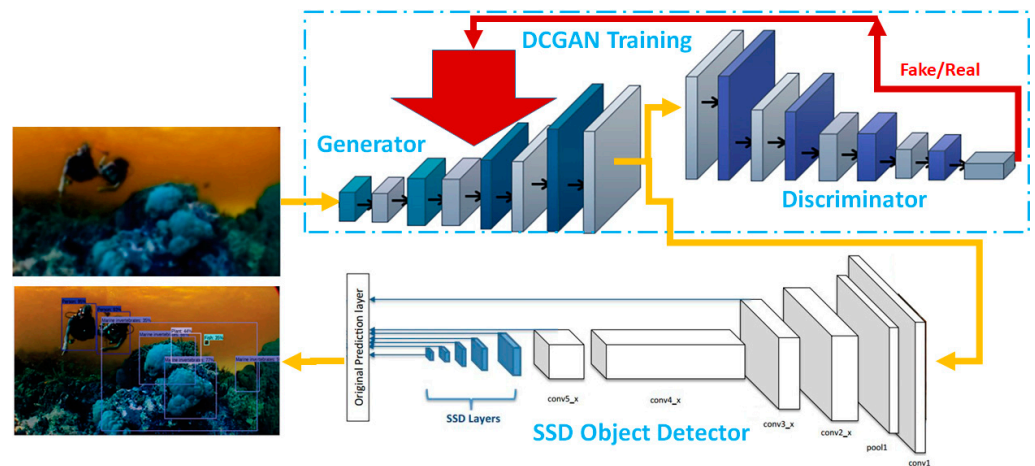
Figure 3. The proposed end-to-end DCGAN+SSD architecture.

In our proposed framework (Figure 3), the DCGAN generator is trained with its discriminator to produce converted images that are more suitable for object detection. After training the DCGAN, the detector is then applied to detect objects from the converted images, instead of directly evaluating the original images. With such a new cascaded architecture, the aim is to overcome many challenges in object detection caused by hostile conditions such as low image quality and unbalanced datasets.

Here, the loss function of the generator was adjusted to suit our model and application, which can be expressed as

$$L_G = \alpha L_C + (1 - \alpha) \times L_{Adv}, \tag{1}$$

where α is the weight combining both loss costs, L_{Adv} is the adversary loss, and L_C is the content loss between the generated image and the original image, which can be computed as the normalization difference shown in the equation below.

$$L_C = \left\| y^{generated} - y^{original} \right\|^2. \tag{2}$$

L_{Adv} is the adversarial loss, as described in the DCGAN [12,13],

$$L_{Adv} = \sum_{n=1}^{N} -\log D(G(y^{input})). \tag{3}$$

The loss of the discriminator through adversarial loss can be expressed as,

$$L_D = -L_{Adv} = \sum_{n=1}^{N} (\log D\left(G\left(y^{input}\right)\right) + \log\left(1 - D\left(y^{original}\right)\right)), \tag{4}$$

where the first term is the loss over generated images from the generator, and the second term is the loss over the original high-quality images.

Among various deep learning-based object detectors, YOLO [16] and SSD [14] are the two state-of-the-art methods that can quickly capture object regions. We chose SSD as our object detector in this work because we are more concerned with accuracy instead of speed for AUVs. The proposed combination is aimed at improving the detection rates of diverse objects in subsea images or videos by simply taking advantage of the DCGAN generator to produce higher-resolution features of the objects.

The pipeline of the whole architecture (Figure 3) can be described as follows: first, the input low-quality images are fed into the generator to produce high-quality images, which are then input to the discriminator in the training process. In the test process, the generated quality-improved features are fed to the detector. The SSD, which was employed in this work, was trained using ImageNet. Notably, the SSD detector performance lags in the scale

factor, whereas the DCGAN can improvise the scale factor in SSD by providing the detector with enhanced images. In the convolutional detector, each feature layer can produce a fixed set of predictions for detection using convolutional filters. The SSD is associated with a set of bounding boxes with default sizes related to each cell from each feature map. The default size of the bounding boxes is extracted in a convolutional manner using the feature map. The position of each bounding box is related to its size in the cell, and the class indicates the presence of the feature cell in the cell bounding boxes.

4. Our PSO-Based Model Optimization

4.1. Particle Swarm Optimization

PSO [37] is an optimization concept for nonlinear functions. Its evolution and characterization compared with other optimization algorithms have shown that PSO outperforms all other optimization algorithms [37]. PSO is related to many methods in artificial intelligence. It is a common methodology, typically applied to bird flocking, fish schooling, and swarming theory [38]. The mechanism of PSO is based on five principles: (1) response with proximity in space and time computations; (2) response to quality factors in the environment; (3) diversity with multiple factors related to the population-based model; (4) stability in selecting the hyperparameters, whereby the populated hyperparameters should not change their behavior as a function of the environment; (5) adaptability based on the data.

The PSO practically uses particles as a population [37], where the m-dimensional vector articulates each particle. Each particle of the PSO is an interpretation of a solution in the m-dimensional search space. The PSO is considered a group of particles initialized in a random manner, before populating the search space. The PSO direction is influenced by two factors, position and velocity, according to the individual best from the previous iteration $p_{best,i}$ and all individual best particles in the previous best swarm, g_{i_best}.

$$p_i^{t+1} = p_i^t + v_i^{t+1}, \tag{5}$$

where,

$$v_i^{t+1} = W v_i^t + c_1 r_1 + (p_{best,i} - x_i) + c_2 r_2 (g_{best,i} - x_i), \tag{6}$$

where W denotes the inertial weights, r denotes the previous position of particles, c denotes the current place, t and $t + 1$ denote the iterations and generations of the new position, i denotes the dimension of the particle, x_i^t denotes the particle in generation t of the position in the dimension i, v_i^{t+1} is the velocity of the particle generation in $t + 1$ and the i dimension.

The hyperparameters respond to the quality factors p_{i_best} and g_{i_best}, as shown in Equations (5) and (6). The allocation of quality factors ensures a diverse response. The parameters change only when g_{i_best} is changed as the best hyperparameter, by observing the principles of stability and adaptability [37,38].

4.2. Hyperparameter Tuning

The challenge in optimizing the parameters is the same in DNN and DCGAN. This work presents a powerful optimization approach using challenging hyperparameters with PSO. Using PSO with a large-scale network like DCGAN+SSD is challenging, and its efficiency depends on hyperparameters such as initial learning rate, decay, and momentum. PSO can be used to improve the performance of the DCGAN+SSD architecture.

Hyperparameter algorithms show better performance than human efforts with trial-and-error experiments [38]. However, we encountered more challenges during implementation in practice when integrating into a large network (DCGAN+SSD), which has very high computational complexity. Few algorithms are very efficient in achieving the target in optimizing problems. In the presented work, we validated PSO for the optimization problem with DCGAN+SSD, performed using the CIFAR-100 dataset. The output results proved that PSO provided an improved output with the existing architecture. They also demonstrated that PSO could efficiently solve the hyperparameters.

Thus, in our work, we handled the problem of hyperparameter optimization for the combination DCGAN+SSD using the PSO algorithm, to improve the classification process in terms of both robustness and fairness toward all targeted classes in the object detection task. The pseudocode of the PSO-based algorithm, i.e., PSO+DCGAN+SSD, is shown in Algorithm 1. This hyperparameter selection process plays a vital role in training the SSD for object detection applications. It is not limited to one network but was applied to DCGAN+SSD in this study. Although both networks could be trained together, we used PSO only on SSD to reduce the complexity. In our work, the DCGAN+SSD network with PSO was customized within an integrated network, while addressing issues such as parallel processing and stability in training.

Algorithm 1. **The Pseudo Code of PSO-Based Optimization**.

Initialize the swarm and the search parameters.
While (K < the maximum number of iterations)
  For (i = 1 to number of particles N)
   {
      Evaluate particle i;
      If the fitness of x_i^t is greater than the fitness of p_{i_best},
      Then update $p_{i\_best} = x_i^t$;
      If the fitness of x_i^t is greater than that of the global best which is g_{best},
       Then update $g_{i\_best} = x_i^t$;
       For (each dimension, i.e., d_1, d_2, and d_3 = dimensions for the learning rate, momentum, and weight decay, respectively),
        Then update velocity vector v_i^{t+1} using the defined PSO velocity updating equation;
        Update particle position x_i^{t+1} using the defined PSO position updating equation;
       End (for dimensions)
   }
  End For Loop (for particles)
End While Loop

To resolve the problems and increase the performance of DCGAN+SSD for object detection tasks, our PSO-based approach mainly addressed the challenge in finding the best hyperparameters. Through PSO-based hyperparameter optimization, the best hyperparameters were identified and the tuning was automated for the object detection model. These optimized parameters (learning rate, decay, and momentum) were then taken to train the model.

This work presents an overview of the PSO with a simplistic layout describing hyperparameter optimization. To determine the best hyperparameters, swarm intelligence is a good candidate, as inspired by the real-world example of birds searching for their food by continuously changing their position until attaining their target. PSO, as indicated by its name, uses particles to find the swarm position, as introduced in [39]. PSO is a heuristic optimization technique, whereby it populates and then updates the state of individuals from the population through an advancement process. When compared to other swarm intelligence algorithms, PSO is simple and easy to handle. Moreover, the solution provided by the PSO algorithm is optimal and tangible.

5. Experimental Results

To compare and differentiate the results obtained from the PSO-optimized DCGAN+SSD and DCGAN+SSD network, the experimental tests were performed on 23 undersea/subsea videos collected for comparison, with an average of 850 frames extracted per video, yielding a total of 19,550 frames. The model was trained using the CIFAR-100, HAR-DAISY, and ROBOFLOW datasets consisting of a total of 100,000 images belonging to 130 classes, where 70,000 images were considered for training. For testing, we used the videos recorded underseas. This implied a big challenge in cross-dataset validation, which is often a realistic situation in real-world applications. In our experiments, we first compared the test results of our DCGAN+SSD method with the SSD only detection, and then we determined how

PSO could further improve the accuracy of the DCGAN+SSD. Our experimental results are detailed below.

5.1. Comparison of DCGAN+SSD with SSD Only

Figure 4 shows the difference between the original images and our DCGAN-enhanced images. Although these images were seemingly similar before and after DCGAN generation, their features were enhanced for object detection. The DCGAN-based approach depends on generative modeling that builds images very closely related to the original images through its generator, and the discriminator co-learns from these generated images by sending feedback to the generator, as illustrated in Figure 3.


(**a**)   (**b**)

Figure 4. The difference in the original images and the images enhanced by DCGAN: (**a**) the original image; (**b**) the image converted by DCGAN.

Here, the "enhanced images" refer to the images converted by DCGAN for improved object detection. Although DCGANs can be used for image super-resolution, a high-

resolution image does not guarantee a better detection rate if its styles or features are not a fit for the detector.

Figure 5 shows the difference in object detection between the SSD only model and our DCGAN+SSD combined model with enhanced object features for the object detection. We can see that only one object was detected by SSD in these four sample images or key frames, as shown in the right column. In comparison, our DCGAN+SSD model, as shown in the right column, could detect most targeted underwater objects.

(**a**)   (**b**)
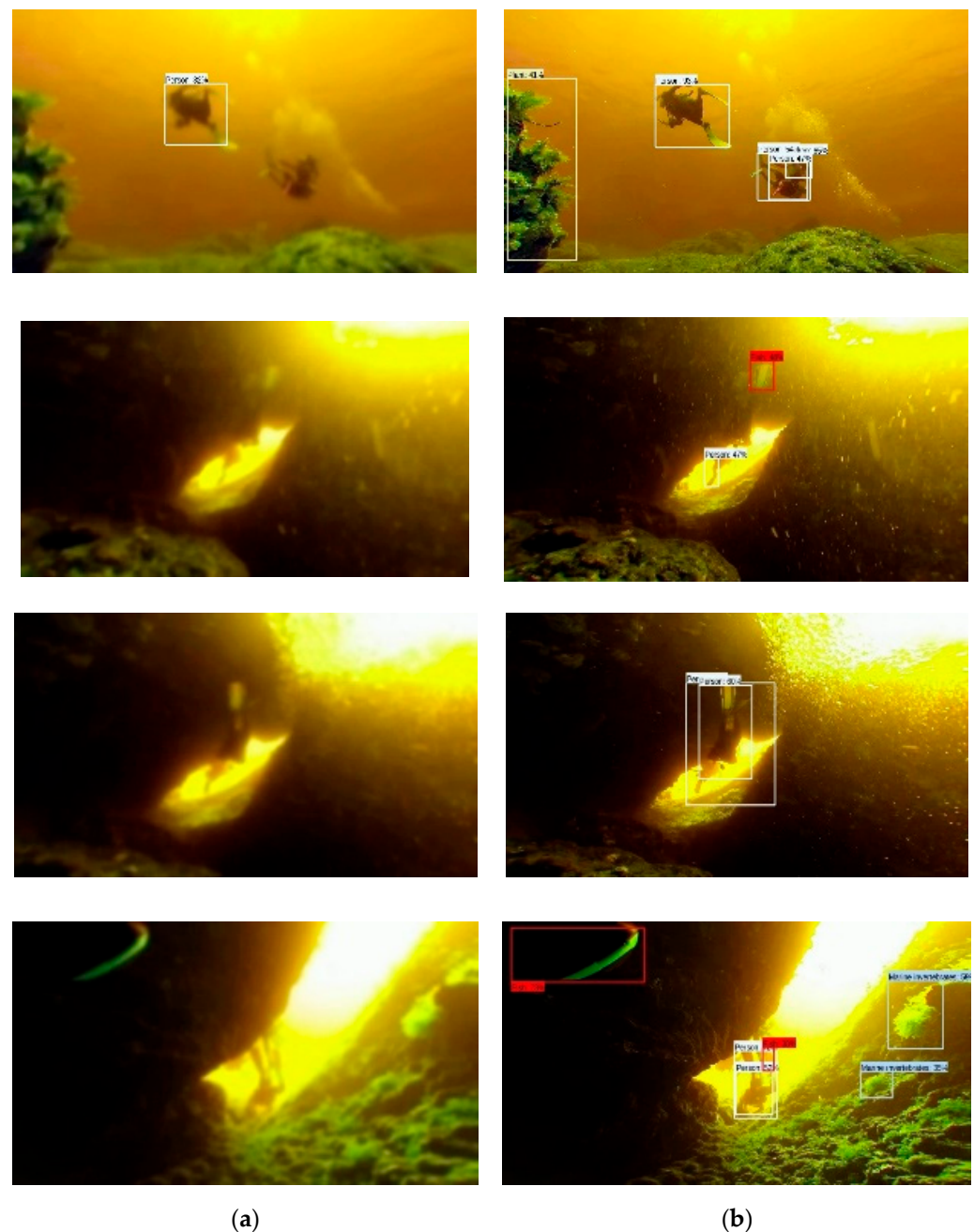
Figure 5. The difference in detection success rates between SSD only and DCGAN+SSD. SSD only missed several objects, while DCGAN+SSD could achieve better detection in all cases: (**a**) object detection by SSD only; (**b**) object detection by DCGAN+SSD.

In our experiment, we chose 230 key frames from the 23 videos at uniform intervals, since most consecutive frames could have been similar to each other in the same videos. Another reason for the reduction in key frames in our tests was to reduce the

effort need to manually label all video frames. As most consecutive frames are similar to each other, our evaluation using key frames could cover most different scenes in these videos. Table 1 shows the results. We can see that the detection was obviously improved by our DCGAN+SSD framework, while the SSD presented a very poor performance on underwater images.

Table 1. Comparison between SSD only and DCGAN+SSD over different categories underwater using 230 key frames.

| Objects	Object Instances	Correct Detection		False Detection	
		SSD	DCGAN-SSD	SSD	DCGAN-SSD
Human	201	5	164	3	12
Fish	62	0	43	0	8
Plants	34	0	14	0	2
Others	13	0	3	0	4

5.2. Our PSO+DCGAN+SSD Model

Object detectors often face loss inefficiency due to incorrect hyperparameters. To improve our model, we introduced PSO to optimize three key hyperparameters, namely, learning rate, momentum, and weight-updating decay. Table 2 shows the hyperparameters from each particle and their fitting loss in the PSO-based hyperparameter optimization. The training was carried out using the CADDY, ROBOFLOW, and HAR-DAISY datasets, consisting of different images with different light conditions and objects. It is worth noting that these datasets are very different from hostile undersea/subsea conditions. As a result of PSO optimization, we can see that the second particle in Table 2 represented the best learning process.

Table 2. Hyperparameters associated with particles in the PSO-based optimization of DCGAN+SSD model.

Learning Rate	Momentum	Weight Decay	Fitness (in Loss)
0.000285	0.007951	0.44710	231.5625
0.006902	**0.00908**	**0.82245**	**9.58052**
0.000521	0.001392	0.2512	198.371
0.001138	0.000235	0.4467	132.3029
0.005217	0.001104	0.92989	16.11632
0.000445	0.003728	0.7033	156.3364
0.003516	0.003442	0.1698	77.77496
0.001108	0.009621	0.5448	122.4384
0.002458	0.007613	0.7595	31.23957
0.001261	0.002016	0.1152	163.3016

One of the main challenges in underwater object detection is background variation. The background varies underwater due to waves, optical diffraction, and various lighting conditions. Another main drawback in underwater object detection is the equipment used, which represents a limitation in most research studies. Obviously, due to data scarcity, we do not have access to large datasets of various hostile undersea conditions, which could be a challenge for our models to address.

Figure 6 shows the difference in object detection tests before and after PSO-based optimization. From the experimental results, it is clear that our proposed PSO+DCGAN+SSD model could robustly detect more underwater objects, in comparison with the initial DCGAN+SSD model before optimization.

(**a**) (**b**)
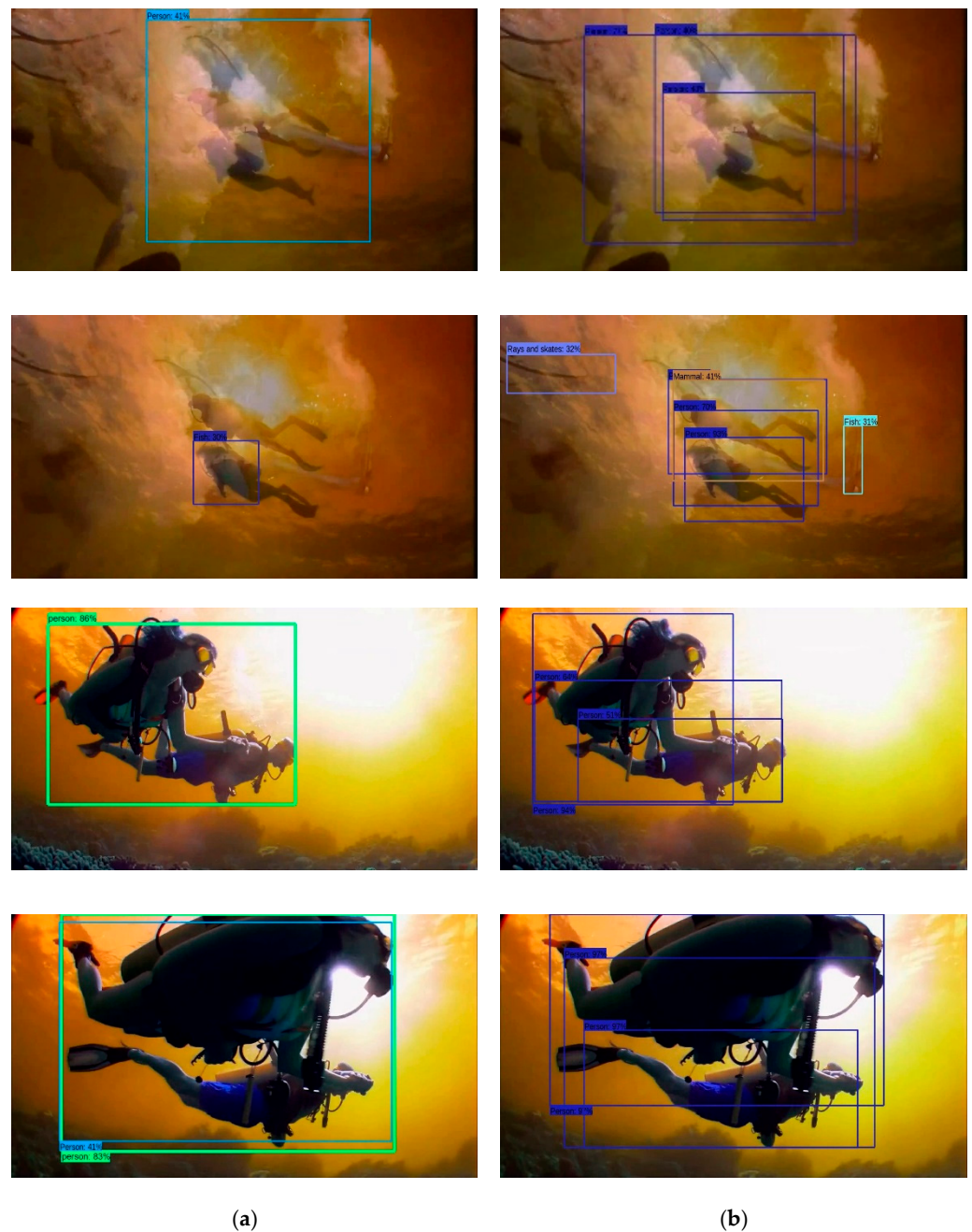
Figure 6. The difference in detection between DCGAN+SSD and PSO+DCGAN+SSD: (**a**) object detection by DCGAN+SSD; (**b**) object detection by PSO+DCGAN+SSD.

Table 3 shows the results over all key frames, revealing that the initial DCGAN+SSD model achieved detection rates of 81.0% for human subjects (divers), 69.3% for undersea fish, 41.0% for undersea plants, and 23.0% for other objects. We can see there was clearly an unbalanced performance that led to a biased model. After PSO optimization, our PSO+DCGAN+SSD model achieved detection rates of 92.5% for human subjects (divers), 93.0% for undersea fish, 70.0% for undersea plants, and 61.5% for other objects, showing a much more robust and fair detection toward different underwater objects.

Table 3. Comparison between DCGAN-SSD and DCGAN-SSD+PSO for different categories underwater.

Categories	Object Instances	Detection		Accuracy	
		DCGAN-SSD	DCGAN-SSD+PSO	DCGAN-SSD	DCGAN-SSD+PSO
Human	201	164	**186**	81.0%	**92.5%**
Fish	62	43	**58**	69.3%	**93.0%**
Plants	34	14	**24**	41.0%	**70.0%**
Others	13	3	**8**	23.0%	**61.5%**

Fairness, as principally defined by the Rawlsian algorithm [29], is a measure to evaluate the merits of the advantaged side over the disadvantaged side. To evaluate how much our PSO-based model could help improve fairness, we used three criteria to evaluate the degree of bias. First, we used the ratio of the max and min accuracy rates for all classes as a criterion to measure the degree of bias of the model.

$$B_r = \frac{A_{max}}{A_{min}} - 1. \tag{7}$$

According to the above formula, if there is no bias, the max and min accuracy rates would be the same, and the ratio bias B_r would be zero, implying no bias.

The second measure used was the absolute difference between the max and min accuracy rates, as defined below.

$$B_d = ||A_{max} - A_{min}||. \tag{8}$$

Similar to the ratio bias, the absolute bias B_d is zero if there is no bias in a model.

The third measure used was the standard deviation of all accuracy rates for all classes, as defined below.

$$B_s = \sqrt{\frac{\sum_{i=1}^{n} \left(A_i - \overline{A}\right)^2}{n-1}}, \tag{9}$$

where

$$\overline{A} = \frac{\sum_{i=1}^{n} ||A_i||}{n}. \tag{10}$$

Similar to the previous bias measures, the standard bias B_s is zero if there is no bias in a model.

Using the above three criteria, we examined the fairness of our models. Figure 7 plots the accuracy rates as a visual comparison between the initial DCGAN+SSD model and the optimized PSO+DCGAN+SSD model. According to the accuracy rates and Equations (7)–(10), we could easily calculate the degrees ratio bias, absolute bias, and standard bias. The initial DCGAN+SSD model attained a ratio bias of 2.52, absolute bias of 0.58, and standard bias of 0.26, while the optimized PSO+DCGAN+SSD model achieved a ratio bias of 0.51, absolute bias of 0.32, and standard bias of 0.16. From the visual comparison shown in Figure 8, we can see that our PSO+DCGAN+SSD model achieved a consistently much lower degree of bias, representing as a robust and fair model in the experiments for all classes according to all three criteria of fairness.
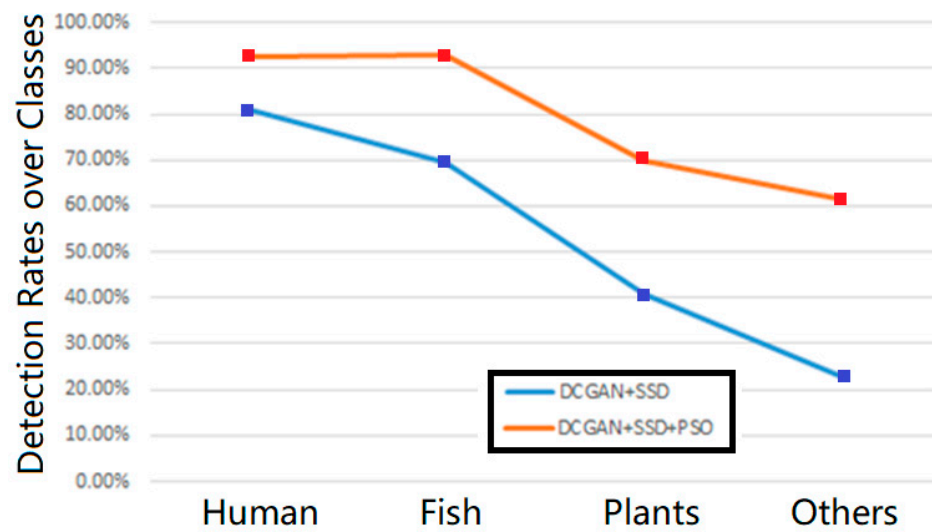
Figure 7. The visual comparison of accuracy rates between the initial DCGAN+SSD model and the optimized PSO+DCGAN+SSD model.
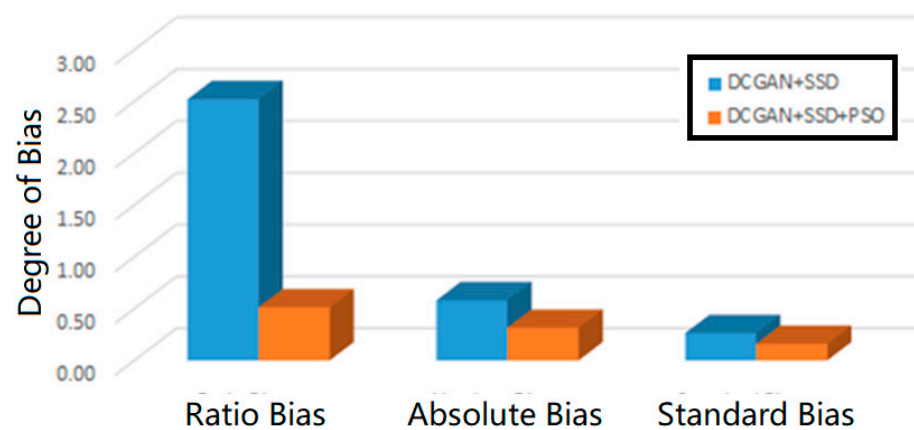
Figure 8. The comparison between the initial DCGAN+SSD model and the optimized PSO+DCGAN+SSD model according to the degrees of ratio bias, absolute bias, and standard bias.

In our simulation, we ran our codes on MATLAB over 4CIF videos. We encountered no problems when processing the input videos in real time (30 frames per second). Although the DCGAN was cascaded with SSD, data processing could be performed in a forward pass, with no apparent delay in speed, whereas we can consider all training to have been performed offline.

In our work, we focused on using underwater cameras. It is worth mentioning that there are various approaches to underwater imaging, such as sonar-based undersea imaging, which has very different features and purposes [40]. In our future work, we will consider these modalities in our AUV systems and exploit these available technologies in a multimodal way.

6. Conclusions

In conclusion, our work targeted the challenges of undersea and subsea object detection from mounted cameras on automated underwater vehicles (AUVs), aiming to offer a powerful platform for subsea/undersea biodiversity studies by addressing the hostile underwater lighting conditions and the data scarcity on samples of rare species. We examined our proposed cascaded model DCGAN+SSD on the basis of the assumption that the DCGAN could help to convert the input images into a more suitable style for the detector. Our experiments successfully testified our assumption and demonstrated that

DCGAN+SSD could handle hostile underwater conditions and achieve better detection rates. Subsequently, we further introduced a PSO-based optimization algorithm to tune the hyperparameters in the DCGAN+SSD model. The experiments proved that such an optimization process could produce a much more robust PSO+DCGAN+SSD model toward underwater object detection. To evaluate the fairness of our models, according to the well-known Rawlsian definition of fairness, we proposed three criteria to measure the degree of bias, namely, ratio bias, absolute bias, and standard bias. Our results clearly showed that the optimized PSO+DCGAN+SSD model not only demonstrated a robust performance in object detection, but also achieved a much better fairness for different objects under hostile underwater conditions, revealing our model as a robust and fair solution for tackling the challenges faced by AUVs.

It is worth noting that, though we validated our assumption that the DCGAN+SSD model could work better than the SSD only model, its underlying theoretical mechanism was not clearly identified. A possible reason for the improved performance is that DCGAN could convert underwater images into styles with features better matching the detector, while both DCGAN and SSD were trained by similar datasets. Such a style conversion could then help overcome the challenges of both the data scarcity of underwater sample images and the low-quality features in underwater conditions. Therefore, although we successfully validated our assumption that DCGAN+SSD could work better than the detector only case for underwater object detection, it is necessary to carry out a deeper theoretical investigation to address this hypothesis, which will be considered in our future work.

Author Contributions: Conceptualization, L.Z., C.-T.L., A.B. and R.J.; methodology, R.D., L.Z. and R.J.; software, R.D.; validation, R.D.; writing—original draft preparation, R.D.; writing—review and editing, all authors. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Data used in this paper were collected from various online public sources including YouTube and Kaggle. No new data was produced from this research. The collected data are available upon request from the corresponding author (drk.ranjithkumar@gmail.com).

Code Availability: The codes developed in this paper are available upon request from the corresponding author (drk.ranjithkumar@gmail.com).

References

1. Moniruzzaman, M.; Islam, S.M.S.; Bennamoun, M.; Lavery, P. Deep Learning on Underwater Marine Object Detection: A Survey. In *Advanced Concepts for Intelligent Vision Systems. ACIVS 2017*; Blanc-Talon, J., Penne, R., Philips, W., Popescu, D., Scheunders, P., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2017; Volume 10617.
2. Costello, C.; Cao, L.; Gelcich, S.; Cisneros-Mata, M.Á.; Free, C.M.; Froehlich, H.E.; Golden, C.D.; Ishimura, G.; Maier, J.; Macadam-Somer, I.; et al. The future of food from the sea. *Nature* **2020**, *588*, 95–100. [CrossRef] [PubMed]
3. Arrow, K.J. Global Climate Change: A Challenge to Policy. *Econ. Voice* **2007**, *4*. [CrossRef]
4. Wunderling, N.; Willeit, M.; Donges, J.F.; Winkelmann, R. Global warming due to loss of large ice masses and Arctic summer sea ice. *Nat. Commun.* **2020**, *11*, 5177. [CrossRef]
5. Kazour, M.; Terki, S.; Rabhi, K.; Jemaa, S.; Khalaf, G.; Amara, R. Sources of microplastics pollution in the marine environment: Importance of wastewater treatment plant and coastal landfill. *Mar. Pollut. Bull.* **2019**, *146*, 608–618. [CrossRef] [PubMed]
6. Borowitzka, M.A. Intertidal algal species diversity and the effect of pollution. *Mar. Freshw. Res.* **1972**, *23*, 73–74. [CrossRef]
7. Strachan, N.J.C. Recognition of fish species by colour and shape. *Image Vis. Comput.* **1993**, *11*, 2–10. [CrossRef]
8. Sands, T. Development of Deterministic Artificial Intelligence for Unmanned Underwater Vehicles (UUV). *J. Mar. Sci. Eng.* **2020**, *8*, 578. [CrossRef]
9. Shirakura, N.; Kiyokawa, T.; Kumamoto, H.; Takamatsu, J.; Ogasawara, T. Collection of Marine Debris by Jointly Using UAV-UUV with GUI for Simple Operation. *IEEE Access* **2021**, *9*, 67432–67443. [CrossRef]
10. Yan, Z.; Zhang, J.; Tang, J. Modified whale optimization algorithm for underwater image matching in a UUV vision system. *Multimed. Tools Appl.* **2021**, *80*, 187–213. [CrossRef]

11. Dinakaran, R.; Sexton, G.; Şeker, H.; Bouridane, A.; Jiang, R. Image resolution impact analysis on pedestrian detection in smart cities surveillance. In Proceedings of the 1st International Conference on Internet of Things and Machine Learning, Liverpool, UK, 17–18 October 2017; pp. 1–8.

12. Dinakaran, R.; Easom, P.; Zhang, L.; Bouridane, A.; Jiang, R.; Edirisinghe, E. Distant Pedestrian Detection in the Wild using Single Shot Detector with Deep Convolutional Generative Adversarial Networks. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14–19 July 2019.

13. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. In Proceedings of the International Conference on Learning Representations (ICLR), San Juan, Puerto Rico, 2–4 May 2016; pp. 1–16.

14. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. SSD: Single shot multibox detector. In *Computer Vision—ECCV 2016*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2016; Volume 9905.

15. Athira, M.V.; Khan, D.M. Recent Trends on Object Detection and image classification: A review. In Proceedings of the 2020 International Conference on Computational Performance Evaluation (ComPE), Shillong, India, 2–4 July 2020.

16. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.

17. Chiang, C.; Barnes, C.; Angelov, P.; Jiang, R. Deep Learning-based Automated Forest Health Diagnosis from Aerial Images. *IEEE Access* **2020**, *8*, 144064–144076. [CrossRef]

18. Jiang, R.; Chazot, P.; Pavese, N.; Crookes, D.; Bouridane, A.; Celebi, M.E. Private Facial Prediagnosis as an Edge Service for Parkinson's DBS Treatment Valuation. *IEEE J. Biomed. Health Inform.* **2022**, *26*, 2703–2713. [CrossRef]

19. Storey, G.; Jiang, R.; Keogh, S.; Bouridane, A.; Li, C.-T. 3DPalsyNet: A Facial Palsy Grading and Motion Recognition Framework Using Fully 3D Convolutional Neural Networks. *IEEE Access* **2019**, *7*, 121655–121664. [CrossRef]

20. Storey, G.; Jiang, R.; Bouridane, A. Role for 2D image generated 3D face models in the rehabilitation of facial palsy. *IET Healthc. Technol. Lett.* **2017**, *4*, 145–148. [CrossRef]

21. Jiang, R.; Crookes, D. Shallow Unorganized Neural Networks Using Smart Neuron Model for Visual Perception. *IEEE Access* **2019**, *7*, 152701–152714. [CrossRef]

22. Jiang, Z.; Chazot, P.L.; Celebi, M.E.; Crookes, D.; Jiang, R. Social Behavioral Phenotyping of Drosophila With a 2D–3D Hybrid CNN Framework. *IEEE Access* **2019**, *7*, 67972–67982. [CrossRef]

23. Jiang, R.; Parry, M.L.; Legg, P.A.; Chung, D.H.S.; Griffiths, I.W. Automated 3-D Animation from Snooker Videos with Information-Theoretical Optimization. *IEEE Trans. Comput. Intell. AI Games* **2013**, *5*, 337–345. [CrossRef]

24. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolution. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015.

25. Shuyang Du, Haoli Guo, Andrew Simpson, "Self-Driving Car Steering Angle Prediction Based on Image Recognition. *arXiv* **2019**, arXiv:1912.05440.

26. Ayachi, R.; Said, Y.; Ben Abdelaali, A. Pedestrian Detection Based on Light-Weighted Separable Convolution for Advanced Driver Assistance Systems. *Neural Process. Lett.* **2020**, *52*, 2655–2668. [CrossRef]

27. Zhao, J.; Mathieu, M.; Goroshin, R.; Lecun, Y. Stacked what-where auto-encoders. *arXiv* **2016**, arXiv:1506.02351.

28. Wang, K.; Liu, M.Z. Object Recognition at Night Scene Based on DCGAN and Faster R-CNN. *IEEE Access* **2020**, *8*, 193168–193182. [CrossRef]

29. Joseph, M.; Kearns, M.; Morgenstern, J.; Neel, S.; Roth, A. Rawlsian Fairness for Machine Learning. In Proceedings of the 3rd Workshop on Fairness, Accountability, and Transparency in Machine Learning, Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD), FATML 2016, New York, NY, USA, 18 November 2016.

30. Bian, Y.; Wang, J.; Jun, J.J.; Xie, X.-Q. Deep Convolutional Generative Adversarial Network (dcGAN) Models for Screening and Design of Small Molecules Targeting Cannabinoid Receptors. *Mol. Pharm.* **2019**, *16*, 4451–4460. [CrossRef] [PubMed]

31. Zhang, J.; Chen, L.; Zhuo, L.; Liang, X.; Li, J. An Efficient Hyperspectral Image Retrieval Method: Deep Spectral-Spatial Feature Extraction with DCGAN and Dimensionality Reduction Using t-SNE-Based NM Hashing. *Remote Sens.* **2018**, *10*, 271. [CrossRef]

32. Dewi, C.; Chen, R.-C.; Liu, Y.-T.; Tai, S.-K. Synthetic Data generation using DCGAN for improved traffic sign recognition. *Neural Comput. Appl.* **2021**, 1–16. [CrossRef]

33. Chen, G.; Liu, L.; Hu, W.; Pan, Z. Semi-Supervised Object Detection in Remote Sensing Images Using Generative Adversarial Networks. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 2503–2506.

34. Wu, Q.; Chen, Y.; Meng, J. DCGAN-Based Data Augmentation for Tomato Leaf Disease Identification. *IEEE Access* **2020**, *8*, 98716–98728. [CrossRef]

35. Kim, B.; Yuvaraj, N.; Preethaa, K.R.S.; Santhosh, R.; Sabari, A. Enhanced pedestrian detection using optimized deep convolution neural network for smart building surveillance. *Soft Comput.* **2020**, *24*, 17081–17092. [CrossRef]

36. Li, Z.; Jin, Y.; Li, Y.; Lin, Z.; Wang, S. Imbalanced Adversarial Learning for Weather Image Generation and Classification. In Proceedings of the 2018 14th IEEE International Conference on Signal Processing (ICSP), Beijing, China, 2–16 August 2018; pp. 1093–1097.

37. Eberhart, R.; Kennedy, J. A New Optimizer using particle swarm theory. In Proceedings of the Sixth International Symposium on Micro Machine and Human Science (MHS'95), Nagoya, Japan, 4–6 October 1995.

38. Bergstra, J.; Bardenet, R.; Bengio, Y.; Kegl, B. Algorithms for Hyper-Parameter Optimization. In Proceedings of the Advances in Neural Information Processing Systems 24 (NIPS 2011), Granada, Spain, 12–17 December 2011.

39. Fukushima, K. Neocognitron: A self-organizing Neural Network Model for Mechanism of Pattern Recognition Unaffected by shift in position. *Biol. Cybern.* **1980**, *36*, 193–202. [CrossRef] [PubMed]

40. Teng, B.; Zhao, H. Underwater target recognition methods based on the framework of deep learning: A survey. *Int. J. Adv. Robot. Syst.* **2020**, *17*, 1–12. [CrossRef]