Article

Plankton Detection with Adversarial Learning and a Densely Connected Deep Learning Model for Class Imbalanced Distribution

Yan Li [1,2], Jiahong Guo [1,2,3], Xiaomin Guo [1,2,4], Zhiqiang Hu [1,2] and Yu Tian [1,2,*]

1. State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China; liyan1@sia.cn (Y.L.); guojiahong@sia.cn (J.G.); guoxiaomin@sia.cn (X.G.); hzq@sia.cn (Z.H.)
2. Institutes of Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang 110169, China
3. School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing 100049, China
4. School of Automation and Electrical Engineering, Shenyang Ligong University, Shenyang 110159, China
* Correspondence: tiany@sia.cn

Abstract: Detecting and classifying the plankton in situ to analyze the population diversity and abundance is fundamental for the understanding of marine planktonic ecosystem. However, the features of plankton are subtle, and the distribution of different plankton taxa is extremely imbalanced in the real marine environment, both of which limit the detection and classification performance of them while implementing the advanced recognition models, especially for the rare taxa. In this paper, a novel plankton detection strategy is proposed combining with a cycle-consistent adversarial network and a densely connected YOLOV3 model, which not only solves the class imbalanced distribution problem of plankton by augmenting data volume for the rare taxa but also reduces the loss of the features in the plankton detection neural network. The mAP of the proposed plankton detection strategy achieved 97.21% and 97.14%, respectively, under two experimental datasets with a difference in the number of rare taxa, which demonstrated the superior performance of plankton detection comparing with other state-of-the-art models. Especially for the rare taxa, the detection accuracy for each rare taxa is improved by about 4.02% on average under the two experimental datasets. Furthermore, the proposed strategy may have the potential to be deployed into an autonomous underwater vehicle for mobile plankton ecosystem observation.

Keywords: plankton detection; class imbalanced distribution; data augmentation; deep learning; adversarial learning

1. Introduction

As a main component of the marine ecosystem, plankton plays an important role in both the global marine carbon cycle and early warning ahead of natural disasters [1,2]. In addition, the plankton with a high-density distribution will also affect the performance of the detecting sensors such as sonar since the acoustic transmission is impeded. Therefore, the research on the comprehensive understanding of the distribution and abundance of the plankton in the marine environment is a focus issue for both ecologists and engineers.

In the past decades and even now, the core ways of plankton sampling are mainly employing traditional tools such as filters, pumps and nets. Furthermore, the collected samples are investigated manually employing expert knowledge in the laboratory environment. It is evident that there are numerous shortcomings. On one hand, the samples are easy to be destroyed during the sampling and investigation, especially for the fragile gelatinous plankton organisms, which would result in a wrong conclusion. On the other hand, this way of plankton sampling and investigation is labor-intensive and time-consuming.

To overcome these shortcomings, the in situ plankton recorder equipment and detection strategy with high accuracy is an urgent demand.

It was not until the late 1970s, the first computing systems with the capability of automatically measuring the planktonic particles within images were introduced [3–5]. After that, especially in the past 20 years, the types of equipment used to record plankton images in situ and analyze them in the laboratory were rapidly developed, such as Video Plankton Recorder [6], FlowCytobot [7], FlowCam [8] and ZooProcess [9]. With the help of these types of equipment, the image data volume of plankton is accumulated rapidly. Simultaneously, to achieve the in situ plankton detection, a number of studies focus on utilizing the image processing technologies to mine from the immense amounts of collected data [9–12]. Following the rapid development of machine learning and computing hardware, deep neural networks are widely implemented in the field of plankton detection because of their superior capability of feature extraction compared with the traditional methods [13,14]. Inspired by AlexNet and VGGNet, Dai J et al. proposed a convolutional neural network (CNN) named ZooplanktoNet consisted of 11 layers and achieved 93.7% accuracy performance on zooplankton detection [13]. Li X et al. and Py O et al. employed a deep residual network (ResNet) and a deep CNN with a multi-size image sensing module for plankton classification, respectively [15,16]. Shi Z employed an improved YOLOV2 (You Only Look Once V2) model to detect the zooplankton in the holographic image data [17]. Pedraza et al. used CNN for the first time in automatic diatom classification and compared the performance between two state-of-the-art models RCNN (Region CNN) and YOLO [18,19]. Kerr T et al. proposed collaborative deep learning models to detect plankton from collected FlowCam image data to solve the problem of class imbalance [20]. Lee et al. incorporate transfer learning by pre-training CNN with class-normalized data and fine-tuning with original data on an open dataset named WHOI-Plankton, the classification accuracy is increased but there remains a significant problem on the prediction quality in rare taxa [21,22]. Lumini A et al. worked on the fine-tuning and the transfer learning of several renowned deep learning models (AlexNet, GoogleNet, VGG, et al.) to design an ensemble of classifiers for plankton, the performance of their approach outperformed other models, and the accuracy baseline achieved about 95.3% accuracy under the WHOI-Plankton dataset [23].

Usually, there are two ways to improve the accuracy of the plankton detection and classification, one is to enrich the amount of the features and the other is to optimize the detection and classification model to reduce the feature loss. Most studies focus on augmenting the volume of the training dataset by rotating the original image data, changing the brightness and other operations [13–20]. Cheng et al. enriched the features of plankton by combining the features under both Cartesian and Polar coordinate systems, and then employed CNN and support vector machines (SVMs) to train the classification model and classify the taxa of plankton [14]. These data augmentation operations are usually for all taxa data, which means the amount of data for all taxa are augmented proportionally, therefore, these augmentation operations do not solve the problems caused by class imbalanced data. Moreover, the learning capability of the detection and classification model is limited for the features of the rare taxa data if the amount of rare taxa data is relatively small. On the other hand, the structure of the DenseNet model [24] with the advantage of the feature reuse ability was fused into other detection models, such as YOLOV3, which reduced the feature loss in the deep neural network model and was proved effective for subtle feature retention [25,26].

To the best of our knowledge, there are two challenges for plankton detection and classification using a deep neural network. First, the plankton is class imbalanced distributed in the spatiotemporal marine environment, this phenomenon limits the detection performance since the neural network is prone to overfitting during model training [27–29]. Second, a large number of the subtle features of plankton are lost during the features transmitting in the neural network because of the convolution and down-sampling operations, which also limits the capability of detection and classification. Therefore, this paper

aims at these two challenges and proposes a novel detection and classification strategy for imbalanced distributed plankton. The main contributions of this paper are summarized as follows. On one hand, an adversarial neural network named CycleGAN is implemented at the pre-processing stage to generate an amount of fake image data to augment the data volume of the rare taxa, which would improve the learning capability of the neural network to the features of the rare taxa [30]. On the other hand, a densely connected YOLOV3 model is proposed to detect and classify the plankton by adding some dense blocks to replace the down-sampling operations of perception layers, which ensure all the features of plankton could transmit in the neural network during model training [31].

The rest of this paper is organized as follows. In Section 2, the data augmentation method based on the CycleGAN model is introduced after reviewing the original dataset. Furthermore, the basis of the original YOLOV3 model and the proposed densely connected model based on it for plankton detection is addressed in Section 2. Subsequently, The performance evaluation metrics are listed in Section 2, while the experimental results are discussed in Section 3. The conclusions for this paper are provided at last in Section 4.

2. Materials and Methods

2.1. Dataset Description and Augmentation

2.1.1. Dataset Description

A large scale and fine-grained dataset for plankton named WHOI-Plankton are used in this work, which is provided by Woods Hole Oceanographic Institution with an Imaging FlowCytobot (IFCB) to imaging plankton since 2006 [21]. The WHOI-Plankton dataset comprises over 3.4 million expert-labeled images covering 100 taxa. However, the data distribution for each taxa is extremely imbalanced by reviewing the WHOI-Plankton dataset, the most volume of the dataset is concentrated in six rare taxa including Detridus, Leptocylindrus, Dino30, Cylindrotheca, Rhizosolenia and Chaetoceros, the total percentage is up to 85% of the whole dataset. This proves the existence of the phenomenon that the plankton taxa are imbalanced distributed in the actual marine environment on one aspect.

Considering that the CycleGAN network needs a certain amount of data to train before expanding the dataset, the taxa with too little data volume will lead to under-fitting of training and affect the quality of the generated data. Therefore, the rare taxa are randomly selected among the taxa with data volumes between 100 and 200 in the WHOI-plankton dataset. The dominant taxa are randomly selected among the taxa with data volume greater than 400. Several taxa are randomly selected and illustrated as shown in Figure 1.
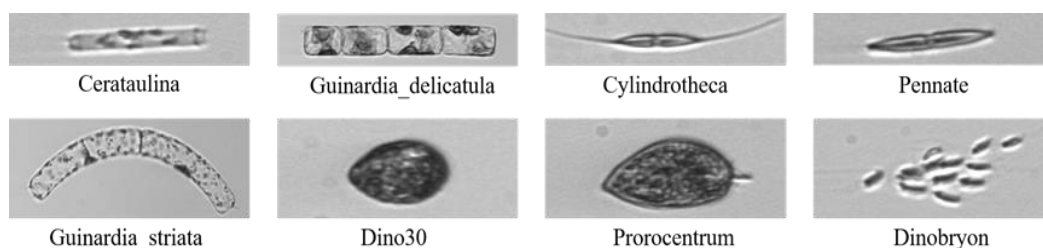

Figure 1. Illustration of the taxa data in this work.

2.1.2. Dataset Augmentation

To improve the learning ability of the detection model to the rare taxa and avoid overfitting during model training, the data volume of the rare taxa is augmented roughly as the same as the dominant taxa before model training. In this paper, a generative adversarial network named CycleGAN is implemented to produce a certain amount of fake data from the unpaired original data to augment the data volume of the rare taxa. The principle of the CycleGAN is shown in Figure 2. The goal is to learn two mapping functions between the domain X and Y ($G : X \rightarrow Y$ and $F : Y \rightarrow X$), and the mapping functions are parameterized by neural networks to fool adversarial discriminators D_Y and

D_X, respectively. These two mapping functions are cycle-consistent, the image x from the domain X should be brought back to the original image by the image transition cycle. Thus, the characteristics of the reconstructed fake images are similar to the original images. The loss function of CycleGAN is formulated as follows:

$$L(G, F, D_X, D_Y) = L_{GAN}(G, D_Y, X, Y) + L_{GAN}(F, D_X, Y, X) + \lambda L_{cyc}(G, F) \qquad (1)$$

where, $L_{GAN}(G, D_Y, X, Y)$ and $L_{GAN}(F, D_X, Y, X)$ are the adversarial loss, $L_{cyc}(G, F)$ is the cycle consistency loss, and λ is a parameter to control the relative importance between marginal matching and cycle consistency. The expectation of CycleGAN is as follows:

$$G^*, F^* = \arg \min_{G,F} \max_{D_X, D_Y} L(G, F, D_X, D_Y) \qquad (2)$$

The detailed mathematical description of CycleGAN also can be found in other literature [29].
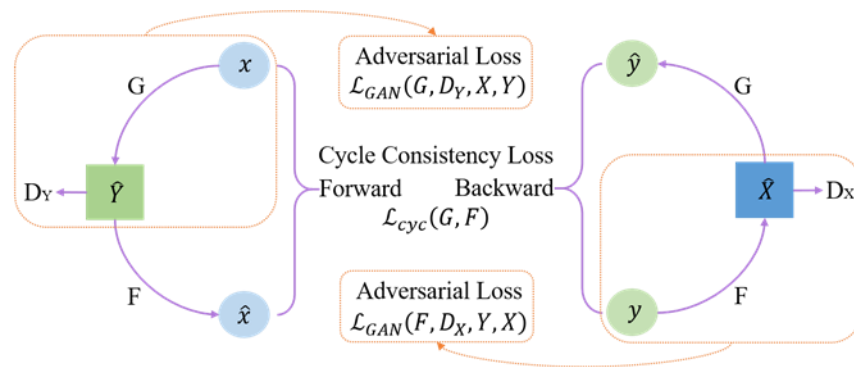

Figure 2. Principle of CycleGAN.

2.2. Plankton Detection Algorithm

2.2.1. Basic of YOLOV3 Model

As a typical one-stage detection model, the YOLO was proposed by Redmon et al. in 2016 [32]. The significant advantage of the YOLO model over the two-stage model based on the region like R-CNN is that it greatly reduces the time consumption of detecting one image [30], which is good for detecting targets in the in situ plankton observation. The basic principle of target detection based on the YOLO model is as follows: the input image is divided into grids. If the center point of the object falls into a grid, the grid is responsible for predicting the object. The prediction bounding box contains five information values: x, y, width, height and prediction confidence. The confidence of the predicted target is defined as follows:

$$Confidence = p_r(Object) \times IoU_{pred}^{truth}, p_r(Object) \in \{0, 1\} \qquad (3)$$

where, the *IoU* is the overlap ratio between the ground truth bounding box and the predicted bounding box. $p_r(Object) = 1$ means the plankton target falls into the grid, and otherwise $p_r(Object) = 0$. Then the dimension of the predicted tensor is as follows:

$$S \times S \times (B * 5 + C) \qquad (4)$$

where, $S \times S$ is the number of grids in the image. B is the number of prediction scales. C is the number of taxa of plankton.

YOLOV3 was first proposed in 2018, which is a classic version of the YOLO series [31]. There are three different prediction scales in the YOLOV3 model with the Darknet-53

structure as a backbone network, which is one of the innovations compared with the previous versions. Therefore, the dimension of the tensor becomes as follows:

$$S \times S \times (3 * (4 + 1 + C)) \tag{5}$$

The loss function of YOLOV3 is composed of coordinate prediction error, *IoU* error and classification error as follows:

$$Loss = \sum_{i=1}^{S^2} Err_{coord} + Err_{IoU} + Err_{cls} \tag{6}$$

where, S^2 is the number of grids in the image.

The coordinate prediction error is defined as follows:

$$\begin{aligned} Err_{coord} = &\lambda_{coord} \sum_{i=1}^{S^2} \sum_{j=1}^{B} I_{ij}^{obj} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] \\ &+ \lambda_{coord} \sum_{i=1}^{S^2} \sum_{j=1}^{B} I_{ij}^{obj} \left[(w_i - \hat{w}_i)^2 + \left(h_i - \hat{h}_i \right)^2 \right] \end{aligned} \tag{7}$$

where, λ_{coord} is the weight of Err_{coord}. $I_{ij}^{obj} = 1$ means the target falls into the *j*th bounding box of the grid *i*, and otherwise $I_{ij}^{obj} = 0$. The four values denote the center coordinates, height and width of the bounding box in (x_i, y_i, w_i, h_i) and $\left(\hat{x}_i, \hat{y}_i, \hat{w}_i, \hat{h}_i \right)$, which means the ground-truth value and the predicted value of the plankton target, respectively.

The *IoU* error is defined as follows:

$$Err_{IoU} = \sum_{i=1}^{S^2} \sum_{j=1}^{B} I_{ij}^{obj} \left(C_i - \hat{C}_i \right)^2 + \lambda_{noobj} \sum_{i=1}^{S^2} \sum_{j=1}^{B} I_{ij}^{noobj} \left(C_i - \hat{C}_i \right)^2 \tag{8}$$

where, λ_{noobj} is the weight of Err_{IoU}, C_i and $\hat{C}_i$ are the true confidence and the predictive confidence of plankton target, respectively.

The classification error is defined as follows:

$$Err_{cls} = \sum_{i=1}^{S^2} \sum_{j=1}^{B} I_{ij}^{obj} \sum_{c \in classes} [\hat{p}_i(c) \log(p_i(c)) + (1 - \hat{p}_i(c)) \log(1 - p_i(c))] \tag{9}$$

where, *c* is the class of the detected target, $p_i(c)$ and $\hat{p}_i(c)$ are the real probability and the prediction probability of the target belonging to the class *c* in the grid *i*, respectively.

2.2.2. Densely Connected Structure

Analysis of the distribution and abundance of rare plankton is a significant part of the investigation of plankton diversity. In order to achieve the purpose, the real-time and accurate identification and classification of plankton become particularly important. This is even more critical in the case of employing mobile underwater vehicles.

Even though the YOLOV3 model has superiority in saving analysis time during detecting plankton targets, the subtle features of plankton are easy to be lost in the process of deepening the neural network layers, which leads to the reduction of the accuracy of plankton identification and classification. The DenseNet was proposed in 2017 with the advantages of promoting feature reuse and reducing gradient disappearance [24], and its structure is shown in Figure 3. In this paper, an improved YOLOV3 model was proposed and it introduced the structure of DenseNet by adding the dense block and transition layer to replace the down-sampling layers of YOLOV3. Therefore, the proposed model ensures the integrity of the feature information in the process of deep neural network propagation.
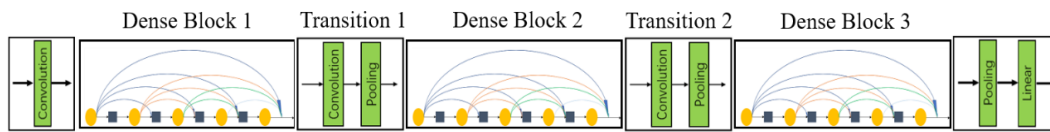
Figure 3. Demonstration of DenseNet structure.

2.2.3. Proposed Plankton Detection Structure

In this paper, the DenseNet structure was integrated into the YOLOV3 model named YOLOV3-dense model proposed to detect the plankton. The purpose of the proposed model is to serve them in situ observation of plankton and mainly based on two advantages as follows. First, the proposed model keeps the lower time cost of the YOLOV3 model and ensures the real-time in situ observation of plankton. Second, the proposed model can better extract the subtle features of plankton and improve detection accuracy. The backbone network structure and the complete network structure of the proposed YOLOV3-dense model are shown in Figures 4 and 5, respectively.

	Type	Filters	Size	Output
	Convolutional	32	3×3	416×416
	Convolutional	64	3×3/2	208×208
1×	Convolutional	32	1×1	
	Convolutional	64	3×3	
	Residual			208×208
	Convolutional	128	3×3/2	104×104
2×	Convolutional	64	1×1	
	Convolutional	128	3×3	
	Residual			104×104
	Convolutional	256	3×3/2	52×52
8×	Convolutional	128	1×1	
	Convolutional	256	3×3	
	Residual			52×52
	Output	256		52×52
	Dense block	*128*	*1×1*	*×8*
	Dense block	*32*	*3×3*	
	Transition	*128*	*1×1*	*×1*
	Transition		*2×2/2*	
	Dense block	*128*	*1×1*	*×8*
	Dense block	*32*	*3×3*	
	DenseNet			*26×26*
	Output	512		26×26
8×	Convolutional	256	1×1	
	Convolutional	512	3×3	
	Residual			26×26
	Output	512		26×26
	Dense block	*128*	*1×1*	*×16*
	Dense block	*32*	*3×3*	
	Transition	*512*	*1×1*	*×1*
	Transition		*2×2/2*	
	Dense block	*128*	*1×1*	*×16*
	Dense block	*32*	*3×3*	
	DenseNet			*13×13*
	Output	1024		13×13
4×	Convolutional	512	1×1	
	Convolutional	1024	3×3	
	Residual			13×13

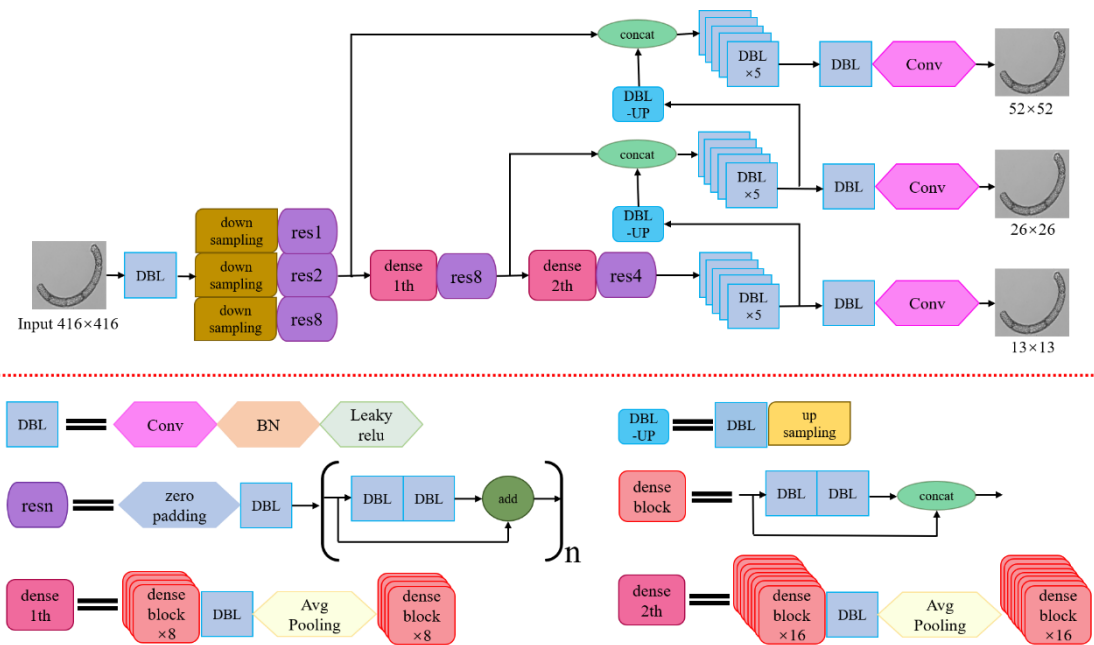**Figure 4.** Network structure diagram of proposed YOLOV3-dense model.

Figure 5. Complete network structure of the proposed YOLOV3-dense model.

In Figures 4 and 5, the input plankton image size is adjusted to 416×416 in prior, and replace the two down-sampling layers (26×26 and 13×13) in YOLOV3 with the DenseNet to avoid the feature loss. The DenseNet structure is composed of the dense-block and the transition layer. The transfer function of the dense block contains three parts, which are Batch Normalization (BN), Rectifying Linear Element (ReLU) and Convolution (Conv), used for nonlinear conversion between $x_0, x_1, \ldots, x_{l-1}$ layers. In the 26×26 down-sampling layer, the input layer x_0 first applies BN-ReLU-Conv (1×1) operation, then applies BN-ReLU-Conv (3×3) operation and output x_1, x_0 and x_1 splicing as the new input $[x_0, x_1]$ and $[x_0, x_1]$ repeats the above operation output x_2. Then the new input becomes $[x_0, x_1, x_2]$, and so on. The transition-layer containing BN-ReLU-Conv (1×1)-average pooling is used to connect adjacent dense blocks. The 13×13 down-sampling layer is the same. Finally, the size of the extracted feature map are $26 \times 26 \times 512$ and $13 \times 13 \times 1024$, respectively, and the feature extraction network outputs three scales feature maps for prediction: 52×52, 26×26 and 13×13.

2.3. Performance Evaluation Metrics

The reasonable index is the favorable basis to evaluate the proposed model. It usually includes detection accuracy and average time cost aspects. For the detection accuracy, precision and recall analysis are utilized to measure it [25,33]. The precision and recall are defined as follows:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{10}$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{11}$$

where, True Positives is the number of targets correctly identified, False Positives is the number of non-targets identified as targets and False Negatives is the number of non-targets identified as non-targets. Therefore, the high precision value means the detection results contain a high percentage of useful information and a low percentage of false alarms. Meanwhile, the higher the recall value is, the larger the proportion of correctly detected targets is.

The average precision (AP) is the integral over the precision-recall curve. In addition, the mean average precision (mAP) is the average precision of all taxa of plankton. These two indexes are defined as follows:

$$AP = \int_0^1 \text{Precision} - \text{Recall}(\text{Recall})d\text{Recall} \tag{12}$$

$$mAP = \frac{1}{C}\sum_{i=1}^{N} AP_i \tag{13}$$

where, C is the number of taxa of plankton.

Furthermore, the average time cost of plankton detection is another important index to evaluate the quality of the proposed model and other comparison models. The lower the average time cost is, the better the real-time performance of the model is and the more practical it is in practical engineering applications.

3. Experiments and Discussions

In order to verify the performance of plankton detection, several well-known and widely used state-of-the-art detection models YOLOV3-tiny, YOLOV3 and Faster RCNN are selected to compare with the proposed YOLOV3-dense model. Table 1 lists some parameters of the proposed model and other comparison models. The proposed detection model and the comparison models in the experiments are performed on a computing server under a Linux environment, which is equipped with Intel XEON Gold 5217 CPU and NVIDIA RTX TITAN GPU cards. A brief flowchart of the experiments is shown in Figure 6.

Table 1. The parameters of proposed model and other comparison models.

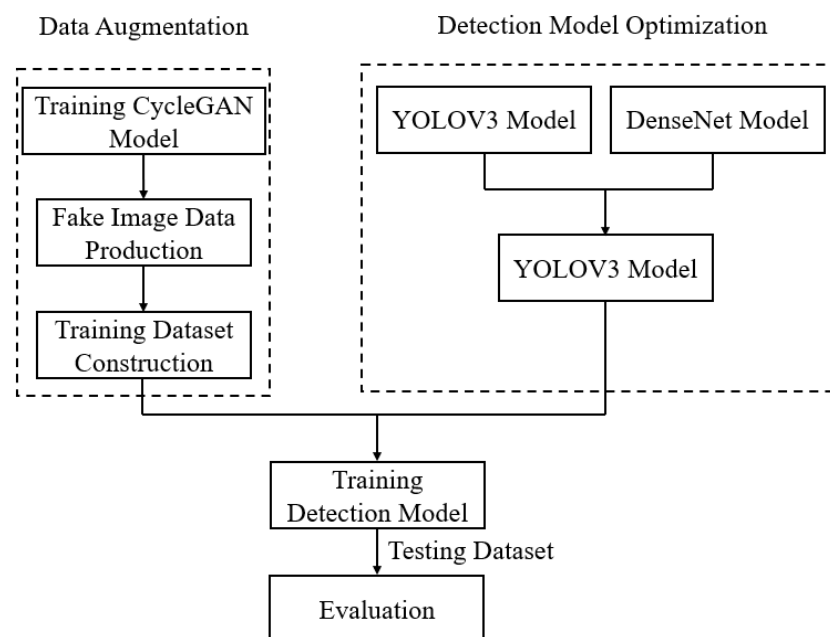| Model	Backbone	Input Size	Boxes	Parameters $\times$ 106
YOLOV3-tiny	Conv-MaxPooling	416×416	2535	8.69
YOLOV3	Darknet-53	416×416	10,647	61.56
YOLOV3-dense	Darknet-dense	416×416	10,647	61.94
Faster-RCNN	ResNet-101	416×416	300	67.66

Figure 6. Flowchart of the experiments.

3.1. Experimental Dataset Production and Components

Both of the original data and the augmented data with the CycleGAN model are labeled manually before training the plankton detection model with a graphical image annotation tool named LabelImg by drawing bounding boxes. Furthermore, the annotated values of plankton are saved as XML files in PASCAL VOC format.

In order to evaluate the performance of the proposed plankton detection strategy to the problem of class imbalanced distribution, one and two taxa are randomly selected as rare taxa to augment the dataset with the CycleGAN model, respectively. The produced fake images of the rare taxa with different training steps are illustrated in Figure 7. It can be seen that the features of the plankton are well learned under the knowledge of humans after training 20,000 steps, and the latter weights achieved are used to produce the fake images and augment them to the training dataset. The components of the dataset with data augmentation for one and two rare taxa are listed in Tables 2 and 3, respectively.
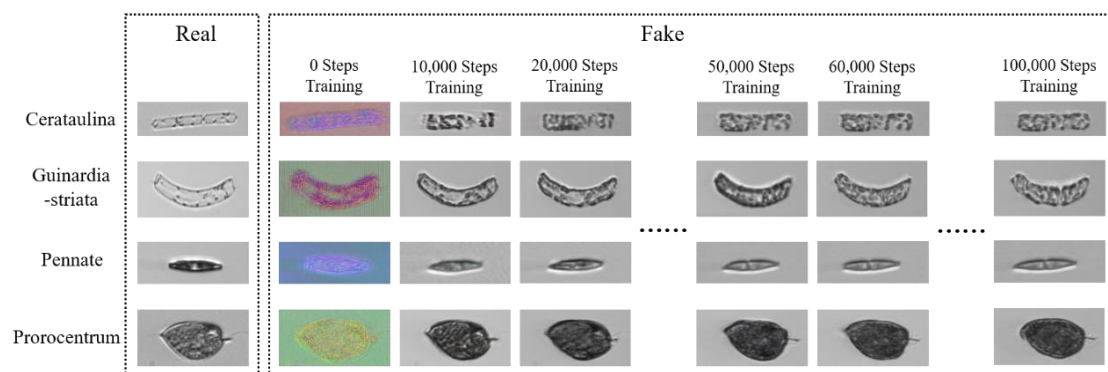

Figure 7. Illustration of fake data production with different training steps.

Table 2. Components of the dataset with data augmentation for one rare taxa.

Taxonomic Group	Training Dataset		Testing Dataset	Total
	Original	Augmentation		
Cerataulina	300	0	100	400
Cylindrotheca	379	0	100	479
Dino30	411	0	100	511
Guinardia_delicatula	450	0	100	550
Guinardia_striata	300	0	100	400
Prorocentrum	60	390	100	550
Total	1900	390	600	2890

Table 3. Components of the dataset with data augmentation for two rare taxa.

Taxonomic Group	Training Dataset		Testing Dataset	Total
	Original	Augmentation		
Cerataulina	300	0	100	400
Cylindrotheca	379	0	100	479
Dino30	411	0	100	511
Dinobryon	348	0	100	448
Guinardia_delicatula	450	0	100	550
Guinardia_striata	300	0	100	400
Pennate	58	362	100	520
Prorocentrum	60	390	100	550
Total	2306	752	800	3858

In Table 2 the taxon "Prorocentrum" is randomly selected as the rare taxa for instance, and the augmented data are produced using different weights of CycleGAN. The data volume of the rare taxon is increased from 60 to 390 after data augmentation which is roughly the same as the volume of the other taxa. The case in Table 3 with data augmentation for two rare taxa is similar. To evaluate the performance of more taxa, other 2 plankton taxa are randomly selected into the experiment and the number of plankton taxa is increased to 8, and another rare taxon "Pennate" with little data volume in the original WHOI-Plankton dataset is added in the experiments. The training data of the rare taxa and all the testing data are strictly and randomly selected from the original WHOI-Plankton dataset considered as ground truth.

3.2. Detection Performance Evaluation

3.2.1. Experiment for the Dataset in Table 2

At the training stages, the loss curves of the YOLOV3 series models are compared with the proposed YOLOV3-dense model, as shown in Figure 8. All of the three YOLOV3 based models achieved convergence after tens of thousands of training steps. The convergence performance of the proposed YOLOV3-dense model is faster than the YOLOV3-tiny model and high degree of consensus as the original YOLOV3 model. The final loss of the original YOLOV3 model, YOLOV3-tiny model and the proposed YOLOV3-dense model is 0.409, 0.514 and 0.405, respectively. This indicates that the proposed YOLOV3-dense model has a higher utilization of image features than the other YOLOV3 based comparison models.
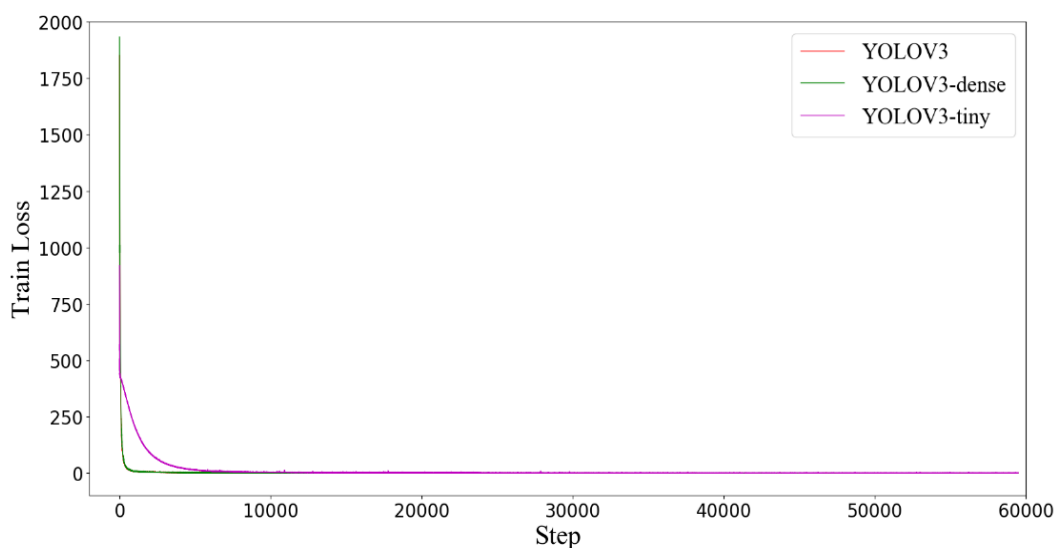

Figure 8. Loss curves of the proposed model and other YOLO series comparison models.

The indexes of performance evaluation for the strategy proposed in this paper and the other comparison models are listed in Table 4. The strategy is abbreviated as ours in the table and the values in bold denote that the related model has the best performance for the corresponding evaluation indexes. Based on the results, the mAP of the proposed strategy achieves 97.21%, which is higher than the other models both YOLOV3 based models and the Faster RCNN model. This verifies the performance of the proposed strategy is superior to the other models in plankton detection. It is notable that the AP of "Prorocentrum" increases from 91.87% to 96.00% after the data augmentation for the rare taxa with the CycleGAN model. Meanwhile, both the true positives and the false positives of the proposed strategy have a better performance than the other comparison models. This indicates that the proposed strategy could detect more plankton accurately with the least false alarms comparing to the other models. On the other hand, another important finding is that all the indexes of performance evaluation for the YOLOV3-dense model

(values for mAP, True Positive and False Positive are 96.55%, 581 and 19, respectively) are better than the YOLOV3 model (values for mAP, True Positive and False Positive are 95.92%, 578 and 22, respectively), which confirms that the densely connected structure is helpful to improve the performance of the plankton detection by reducing the feature loss during the feature transmission in models.

Table 4. Plankton detection performance of the proposed strategy and comparison models for the dataset in Table 2.

Taxonomic Group	Model	YOLOV3-Tiny	YOLOV3	YOLOV3-Dense	Ours	Faster RCNN
AP	Cerataulina	85.63%	94.60%	94.69%	93.54%	86.00%
	Cylindrotheca	98.81%	99.00%	99.00%	99.00%	99.00%
	Dino30	99.50%	99.88%	98.80%	100.00%	99.98%
	Guinardia_delicatula	96.67%	96.00%	97.98%	97.94%	99.66%
	Guinardia_striata	89.76%	97.01%	96.94%	96.75%	99.60%
	Prorocentrum	95.00%	89.00%	91.87%	96.00%	83.00%
mAP		94.23%	95.92%	96.55%	**97.21%**	94.54%
True positives		572	578	581	**584**	568
False positives		28	22	19	**16**	31

3.2.2. Experiment for the Dataset in Table 3

For the dataset in Table 3, the loss curves of the YOLO series models at the training stages are shown in Figure 9, which are most similar to the curves in Figure 8. Even though both the number of taxa and the data volume are increased, the models also could be well trained. The final loss of the original YOLOV3 model, YOLOV3-tiny model and the proposed YOLOV3-dense model is 0.416, 0.469 and 0.423, respectively.
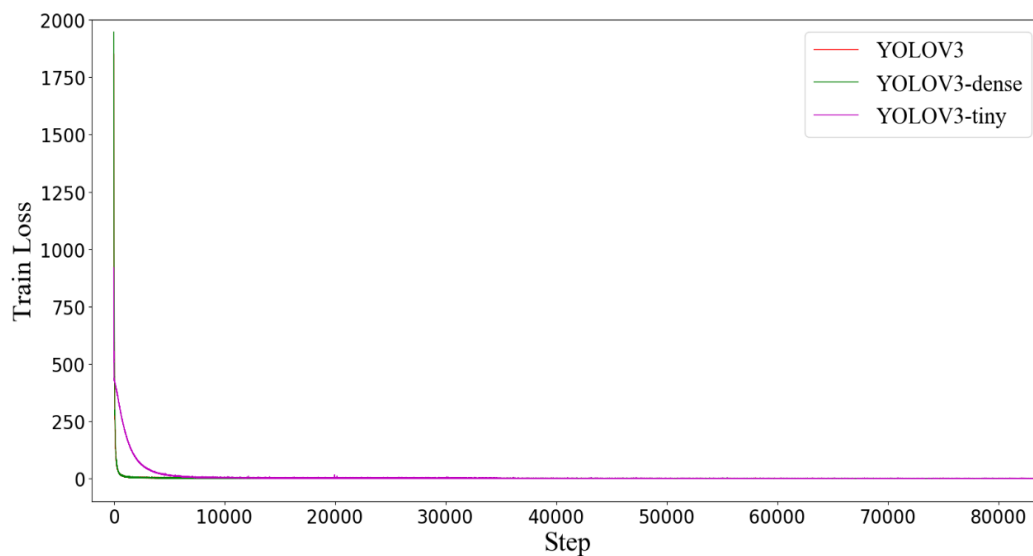

Figure 9. Loss curves of the proposed model and other YOLO series comparison models.

Table 5 lists the indexes of performance evaluation for the dataset in Table 3. Based on the results, the mAP of the YOLOV3-dense model without data augmentation for the rare taxa yields 95.69%, which has 3.02% increase than the YOLOV3-tiny model (92.67%) and basically equals to the YOLOV3 model (95.53%) only has 0.16% increase. However, after the data augmentation for the rare taxa, the mAP of our proposed plankton detection strategy increases to 97.14% with the best detection performance than the other comparison models. Similar to the results with one rare taxon data augmentation, the AP of the two rare taxa are increased from 90.89% to 92.82% and from 92.00% to 98%, respectively. Parallelly,

the true positive and false positive of our proposed detection strategy achieve the best performance than the other comparison models. On the whole, the experimental results for the dataset both in Tables 2 and 3 demonstrate the proposed strategy is suitable for the detection of the imbalanced distributed plankton in the practical ocean environment.

Table 5. Plankton detection performance of the proposed strategy and comparison models for the dataset in Table 3.

Taxonomic Group	Model	YOLOV3-Tiny	YOLOV3	YOLOV3-Dense	Ours	Faster RCNN
AP	Cerataulina	85.13%	92.34%	91.83%	93.27%	82.00%
	Cylindrotheca	96.59%	97.62%	98.88%	98.97%	99.00%
	Dino30	98.58%	99.50%	99.54%	99.73%	99.99%
	Dinobryon	98.93%	99.98%	99.96%	99.88%	100.00%
	Guinardia_delicatula	98.61%	98.76%	97.65%	97.88%	100.00%
	Guinardia_striata	87.48%	98.31%	94.75%	96.57%	97.63%
	Pennate	80.01%	86.76%	90.89%	92.82%	93.76%
	Prorocentrum	96.00%	91.00%	92.00%	98.00%	87.97%
mAP		92.67%	95.53%	95.69%	**97.14%**	95.04%
True positives		753	768	768	**780**	762
False positives		47	32	32	**20**	37

3.3. Real-Time Performance Evaluation

The plankton detection time consumption with these models are listed in Table 6. The average detection time costs of the proposed YOLOV3-dense model are 36 ms and 51 ms for one testing image data in the two experiments, which are slower than the YOLOV3-tiny model and YOLOV3 model, respectively, for the reason that more features were processed and transmitted in the model. Considering the properties of both the data acquisition platform and equipment, the detection speed of YOLOV3-dense is enough for practical applications in real-time. In contrast, the average detection time costs of Faster RCNN are 893 ms and 814 ms, more than 15 times slower than the YOLOV3-dense model. The slow detection speed causes that it is difficult to be implemented in the plankton detection applications with some fast-moving mobile platforms.

Table 6. Comparisons of real-time zooplankton detection performance.

Model	Detection Time Consumption (ms)	
	Dataset in Table 1	Dataset in Table 2
YOLOV3-tiny	8	11
YOLOV3	25	28
YOLOV3-dense	36	51
Faster RCNN	893	814

4. Conclusions

The main goal of this study is to improve the ability of in situ plankton detection for the phenomenon of class imbalanced distribution in the real marine environment. The CycleGAN model was employed to produce many fake images by the adversarial learning and augment the volume of the training dataset for the rare plankton taxa, which ensures the balanced learning of the latter proposed plankton detection model for the features of each plankton taxon. Moreover, an improved plankton detection model based on the YOLOV3 model by fusing the DenseNet was designed, which reduced the feature loss during the transmission in the model.

The experimental results under two experimental datasets with a difference in the number of rare taxa showed that the AP of the rare taxa increases by about 4.02% on average (4.13% for Prorocentrum in Experiment 1; 1.93% and 6% for Pennate and Pennate, respectively, in Experiment 2) and the mAP increases by 0.66%, 1.45%, respectively, after

data augmentation. In addition, the mAP of the proposed model (97.21% in Experiment 1; 97.14% in Experiment 2) outperformed the YOLOV3-tiny, YOLOV3 and Faster-RCNN models (94.23%, 95.92% and 94.54% in Experiment 1; 92.67%, 95.53% and 95.04% in Experiment 2), and the detection time consumption (36 ms in Experiment 1; 51 ms in Experiment 2) is not much different from the YOLOV3-tiny (8 ms in Experiment 1; 11 ms in Experiment 2) and YOLOV3 (25 ms in Experiment 1; 28 ms in Experiment 2) models, but much lower than the Faster-RCNN model (893 ms in Experiment 1; 814 ms in Experiment 2). Hence, the proposed plankton detection strategy in this paper outperformed other state-of-the-art detection models to solve the problem of the species imbalanced distribution both in the performance of accuracy and in real-time.

Currently, the proposed model is deployed on the deep learning development board Jetson Nano which is a small integrated hardware equipped with a Linux system and GPU. The advantage of low energy consumption is helpful to carry out the applications of large-scale to the plankton observation with an underwater autonomous vehicle. In the ongoing and future works, the proposed in situ plankton detection will be implemented on an autonomous underwater vehicle to verify the feasibility in the real marine environment. It is notable that, the autonomous underwater vehicle at higher navigation speed affects the image quality of the imaging sensor which possibly limits the performance of the plankton detection and classification. However, the autonomous underwater vehicle is difficult to control at very low navigation speed in the complex marine environment. Therefore, the plankton sampling strategy and the detection model will be further optimized.

Author Contributions: Conceptualization, Y.L.; methodology, Y.L. and J.G.; software, J.G.; Resources, X.G. and Z.H.; Writing—original draft preparation, Y.L. and J.G.; writing—review and editing, Y.L.; Funding acquisition, Y.T. and Y.L. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data Availability Statement at https://arxiv.org/abs/1510.00745 (accessed on 5 June 2021).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Du, Z.; Xia, C.; Fu, L.; Zhang, N.; Li, B.; Song, J.; Chen, L. A cost-effective in situ zooplankton monitoring system based on novel illumination optimization. *Sensors* **2020**, *20*, 3471. [CrossRef] [PubMed]
2. Tang, X.; Lin, F.; Samson, S.; Remsen, A. Binary plankton image classification. *IEEE J. Ocean. Eng.* **2006**, *31*, 728–735. [CrossRef]
3. Ortner, P.B.; Cummings, S.R.; Aftring, R.P. Silhouette photography of oceanic zooplankton. *Nature* **1979**, *277*, 50–51. [CrossRef]
4. Jeffries, H.P.; Sherman, K.; Maurer, R.; Katsinis, C. Computer-processing of zooplankton samples. In *Estuarine Perspectives*; Academic Press: Cambridge, MA, USA, 1980; pp. 303–316.
5. Rolke, M.; Lenz, J. Size structure analysis of zooplankton samples by means of an automated image analyzing system. *J. Plankton Res.* **1984**, *6*, 637–645. [CrossRef]
6. Davis, C.S.; Gallager, S.M.; Berman, M.S.; Haury, L.R.; Strickler, J.R. The video plankton recorder (VPR): Design and initial results. *Arch. Hydrobiol. Beih* **1992**, *36*, 67–81.
7. Olson, R.J.; Sosik, H.M. A submersible imaging-in-flow instrument to analyze nano-and microplankton: Imaging FlowCytobot. *Limnol. Oceanogr. Methods* **2007**, *5*, 195–203. [CrossRef]
8. Sieracki, C.K.; Sieracki, M.E.; Yentsch, C.S. An imaging-in-flow system for automated analysis of marine microplankton. *Mar. Ecol. Prog. Ser.* **1998**, *168*, 285–296. [CrossRef]

9.　Grosjean, P.; Picheral, M.; Warembourg, C.; Gorsky, G. Enumeration, measurement, and identification of net zooplankton samples using the ZOOSCAN digital imaging system. *ICES J. Mar. Sci.* **2004**, *61*, 518–525. [CrossRef]

10.　Jeffries, H.P.; Berman, M.S.; Poularikas, A.D.; Katsinis, C.; Melas, I.; Sherman, K.; Bivins, L. Automated sizing, counting and identification of zooplankton by pattern recognition. *Mar. Biol.* **1984**, *78*, 329–334. [CrossRef]

11.　Tang, X.; Stewart, W.K.; Huang, H.; Gallager, S.M.; Davis, C.S.; Vincent, L.; Marra, M. Automatic plankton image recognition. *Artif. Intell. Rev.* **1998**, *12*, 177–199. [CrossRef]

12.　Gorsky, G.; Ohman, M.D.; Picheral, M.; Gasparini, S.; Stemmann, L.; Romagnan, J.; Cawood, A.; Pesant, S.; García-Comas, C.; Prejger, F. Digital zooplankton image analysis using the ZooScan integrated system. *J. Plankton Res.* **2010**, *32*, 285–303. [CrossRef]

13.　Dai, J.; Wang, R.; Zheng, H.; Ji, G.; Qiao, X. Zooplanktonet: Deep convolutional network for zooplankton classification. In Proceedings of the OCEANS 2016-Shanghai, Shanghai, China, 10–13 April 2016; IEEE: New York, NY, USA, 2016; pp. 1–6.

14.　Cheng, X.; Ren, Y.; Cheng, K.; Cao, J.; Hao, Q. Method for training convolutional neural networks for in situ plankton image recognition and classification based on the mechanisms of the human eye. *Sensors* **2020**, *20*, 2592. [CrossRef]

15.　Li, X.; Cui, Z. Deep residual networks for plankton classification. In Proceedings of the OCEANS 2016 MTS/IEEE Monterey, Monterey, CA, USA, 19–23 September 2016; IEEE: New York, NY, USA, 2016; pp. 1–4.

16.　Py, O.; Hong, H.; Zhongzhi, S. Plankton classification with deep convolutional neural networks. In Proceedings of the 2016 IEEE Information Technology, Networking, Electronic and Automation Control Conference, Chongqing, China, 20–22 May 2016; IEEE: New York, NY, USA, 2016; pp. 132–136.

17.　Shi, Z.; Wang, K.; Cao, L.; Ren, Y.; Han, Y.; Ma, S. Study on holographic image recognition technology of zooplankton. *DEStech Trans. Comput. Sci. Eng.* **2019**, 580–594. [CrossRef]

18.　Pedraza, A.; Bueno, G.; Deniz, O.; Cristóbal, G.; Blanco, S.; Borrego-Ramos, M. Automated diatom classification (Part B): A deep learning approach. *Appl. Sci.* **2017**, *7*, 460. [CrossRef]

19.　Pedraza, A.; Bueno, G.; Deniz, O.; Ruiz-Santaquiteria, J.; Sanchez, C.; Blanco, S.; Borrego-Ramos, M.; Olenici, A.; Cristobal, G. Lights and pitfalls of convolutional neural networks for diatom identification. In *Optics, Photonics, and Digital Technologies for Imaging Applications V*; International Society for Optics and Photonics: Strasbourg, France, 2018; Volume 10679, p. 106790G.

20.　Kerr, T.; Clark, J.R.; Fileman, E.S.; Widdicombe, C.E.; Pugeault, N. Collaborative deep learning models to handle class imbalance in FlowCam plankton imagery. *IEEE Access* **2020**, *8*, 170013–170032. [CrossRef]

21.　Orenstein, E.C.; Beijbom, O.; Peacock, E.E.; Sosik, H.M. Whoi-plankton-a large scale fine grained visual recognition benchmark dataset for plankton classification. *arXiv* **2015**, arXiv:1510.00745.

22.　Lee, H.; Park, M.; Kim, J. Plankton classification on imbalanced large scale database via convolutional neural networks with transfer learning. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; IEEE: New York, NY, USA, 2016; pp. 3713–3717.

23.　Lumini, A.; Nanni, L. Deep learning and transfer learning features for plankton classification. *Ecol. Inform.* **2019**, *51*, 33–43. [CrossRef]

24.　Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.

25.　Li, Y.; Guo, J.; Guo, X.; Liu, K.; Zhao, W.; Luo, Y.; Wang, Z. A novel target detection method of the unmanned surface vehicle under all-weather conditions with an improved YOLOV3. *Sensors* **2020**, *20*, 4885. [CrossRef]

26.　Tian, Y.; Yang, G.; Wang, Z.; Wang, H.; Li, E.; Liang, Z. Apple detection during different growth stages in orchards using the improved YOLO-V3 model. *Comput. Electron. Agric.* **2019**, *157*, 417–426. [CrossRef]

27.　Cardie, C.; Howe, N. *Improving Minority Class Prediction Using Case-Specific Feature Weights*; Computer Science, Faculty Publications, Smith College: Northampton, MA, USA, 1997.

28.　Haixiang, G.; Yijing, L.; Shang, J.; Mingyun, G.; Yuanyue, H.; Bing, G. Learning from class-imbalanced data: Review of methods and applications. *Expert Syst. Appl.* **2017**, *73*, 220–239. [CrossRef]

29.　Johnson, J.M.; Khoshgoftaar, T.M. The effects of data sampling with deep learning and highly imbalanced big data. *Inf. Syst. Front.* **2020**, *22*, 1113–1131. [CrossRef]

30.　Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.

31.　Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.

32.　Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.

33.　Li, Y.; Xia, C.; Lee, J. Detection of small-sized insect pest in greenhouses based on multifractal analysis. *Opt. Int. J. Light Electron Opt.* **2015**, *126*, 2138–2143. [CrossRef]