Computers and Electronics in Agriculture 185 (2021) 106135

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Real-time detection of uneaten feed pellets in underwater images for aquaculture using an improved YOLO-V4 network

Xuelong Hu [a], Yang Liu [a,b,c,d], Zhengxi Zhao [b,c,d], Jintao Liu [b,c,d,f], Xinting Yang [b,c,d], Chuanheng Sun [b,c,d], Shuhan Chen [a], Bin Li [e,*], Chao Zhou [b,c,d,*]

[a] School of Information Engineering, Yangzhou University, Yangzhou 225127, China
[b] National Engineering Research Center for Information Technology in Agriculture, Beijing 100097, China
[c] Beijing Research Center for Information Technology in Agriculture, Beijing 100097, China
[d] National Engineering Laboratory for Agri-product Quality Traceability, Beijing 100097, China
[e] Beijing Research Center of Intelligent Equipment for Agriculture, Beijing 100097, China
[f] University of Almería, Almería 04120, Spain

ARTICLE INFO

Keywords:
Aquaculture
Improved YOLO-V4 network
Underwater object detection
Uneaten feed pellets
Deep learning

ABSTRACT

In aquaculture, the real-time detection and monitoring of feed pellet consumption is an important basis for formulating scientific feeding strategies that can effectively reduce feed waste and water pollution, which is a win-win scenario in terms of economic and ecological benefits. However, low-quality underwater images and extremely small targets present great challenges to feed pellet detection. To overcome these challenges, this paper proposes an uneaten feed pellet detection model using an improved You Only Look Once (YOLO)-V4 network for aquaculture. The specific implementation methods are as follows: (1) The feature map responsible for large-scale information in the original YOLO-V4 network is replaced by a finer-grained YOLO feature map by modifying the connection mode of the feature pyramid network (FPN) + path aggregation network (PANet). (2) The residual connection mode in CSPDarknets is modified via a DenseNet, which further improves the feature reuse and the network performance. (3) Finally, a de-redundancy operation is carried out to reduce the complexity of the YOLO-V4 network while ensuring the detection accuracy. Experimental results in a real fish farm showed that the detection accuracy is better than that of the original YOLO-V4 network, and the average precision is improved from 65.40% to 92.61% (when the intersection over union is 0.5), for an increase of 27.21%. Additionally, the amount of computation is reduced by approximately 30%. Therefore, the improved YOLO-V4 network can effectively detect underwater feed pellets and is applicable in actual aquaculture environments.

1. Introduction

Aquaculture produces two-thirds of the world's aquatic products, which are an important high-quality protein source for humans (Chalamaiah et al., 2012; Wei et al., 2020; Zhou et al., 2018). Moreover, with the increasing demand for high-quality aquatic products, the welfare of fish in the breeding process has also been given ever-increasing attention. However, many factors can affect the welfare of fish. One of the most important factors is unreasonable feeding, which often occurs, especially overfeeding. It has been reported that more than 60% of the feed in the culture system exists in the form of tiny particles (Masser, 1992). The decomposition of these particles consumes oxygen and produces ammonia and other toxic substances, which affect fish welfare and growth (Chen et al., 2020; Yang et al., 2021). Many previous studies clearly showed that an increase in ammonia concentration may cause physiological disturbances in the blood chemistry, osmoregulatory capability, oxygen consumption, oxidative stress, and antioxidant status of farmed fish, leading to changes in feeding behaviors, declines in growth and immunity, or even death (Harsij et al., 2020; Zhang et al., 2019). In addition, feed waste increases the proportion of feed contributions to the total cost (De Verdal et al., 2018); for example, this proportion is 40–50% for Atlantic salmon (Føre et al., 2016; Liu et al., 2014); 60–80% for carp (Føre et al., 2011; Wu et al., 2015); and 86% for catfish (Rola and Hasan, 2007). Therefore, the real-time detection and

* Corresponding authors at: Shuguang Huayuan Middle Road 9#, Haidian District, Beijing 100097, China.
E-mail address: zhouc@nercita.org.cn (C. Zhou).

https://doi.org/10.1016/j.compag.2021.106135
Received 15 September 2020; Received in revised form 20 March 2021; Accepted 25 March 2021
Available online 3 April 2021

monitoring of uneaten feed can effectively reduce the occurrence of overfeeding. Thus, the cost is reduced, and the water quality can be significantly improved, which is of great scientific significance for guiding production.

However, there are many problems in feed detection and monitoring, such as small targets, complex backgrounds and fish interference, which present great challenges in the identification of underwater feed pellets. In previous studies, acoustic technology was used to detect underwater feed pellets. For example, Llorens et al. (2017) quantified the amount of uneaten feed pellets dropped by an ultrasonic echo method. Previously, Juell (1991) used an echo integration method to estimate the falling pellet abundance. Nevertheless, an automated acoustical application needs an acoustical characterization of the pellets and their condition (Simmonds and MacLennan, 2008), such as the backscattered energy per unit volume or the backscattering cross-section of an average single scatterer. Studies by other scholars have shown that although sonar systems can be used at night, their practical applications are limited by their monochrome, low-quality images (Terayama et al., 2019; Zhou et al., 2018). However, acoustic technology is expensive, and it is easily disturbed by noise, which limits its application in practical production (Han et al., 2009). In recent years, machine vision has developed rapidly due to the advantages of nondestructiveness, low cost and ease of development. At the same time, many specific image preprocessing and enhancement algorithms have emerged (Chiang and Chen, 2011; Li et al., 2016; Sun et al., 2020). These factors make it possible to detect pellets through machine vision. Limited by the software and hardware processing level, Foster et al. (1995) observed and calculated the number of uneaten feed pellets with the naked eye by placing an underwater camera in the tank. With the development of image processing techniques, Atoum et al. (2015) used a correlation filter to detect underwater feed pellets. Li et al. (2017) proposed an adaptive threshold segmentation method to detect uneaten fish food in underwater images. The above method is simple and easy to deploy and has achieved good results in practical applications. However, the detection of underwater uneaten feed pellets also faces many challenges, such as image blurring, small-target problems, high pellet density, and motion blur. Moreover, due to the many imaging problems and the difficulty of underwater target detection, traditional image processing methods require manual selection of the features of the detection target, such as the shape and circumference, and the final accuracy is greatly affected by the selected features. Therefore, the accuracy of the detection of uneaten feed pellets in real scenes still needs to be improved.

In recent years, deep learning technology has developed, and it has achieved far higher accuracy than traditional machine learning because it can automatically extract high-dimensional features from massive information (LeCun et al., 2015). It has been widely used in fish identification, behavior analysis and water quality prediction in the early stage (Yang et al., 2021). For example, Rauf et al. (2019) used deep convolutional neural networks to automatically identify fish based on visual features; Måløy et al. (2019) proposed a dual-stream recurrent network (DSRN) to automatically capture the spatiotemporal behavior of salmon during swimming; Fernandes et al. (2020) proposed deep learning image segmentation for the extraction of fish body measurements and prediction of body weight and carcass traits in Nile tilapia; Zhang et al. (2020) performed automatic fish population counting by using machine vision and a hybrid deep neural network model; and Labao and Naval (2019) proposed a fish detection system based on deep network architectures to robustly detect and count fish targets under a variety of benthic background and illumination conditions. Similarly, the underwater feed pellet detection task has unique and complex features, and both speed and high precision are needed. Regarding the accuracy of feed pellet detection, due to the tiny targets, the relative error of the detection algorithm in predicting the position information will be greater than that for a large target. Similarly, the occlusion of two or more small targets is easier to predict as one single target. In actual production, the detection speed is an important factor that cannot be ignored (Zou et al., 2019). Uneaten feed pellet monitoring requires high real-time performance, and a time lag will not be conducive to subsequent feedback control.

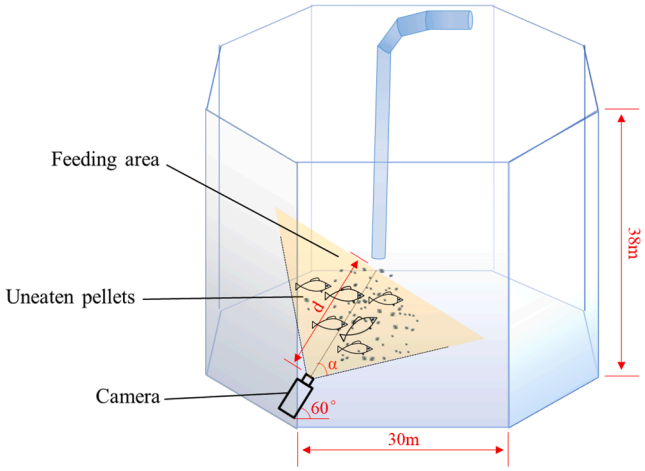


Fig. 1. Cage dimensions and image acquisition schematic.

As a current excellent one-stage detection algorithm, the You Only Look Once (YOLO) series algorithm has high detection accuracy and fast detection speed and is widely used in various target detection tasks. For example, Tian et al. (2019) used improved YOLO-V3 to detect apples at different growth stages; Shi et al. (2020) proposed a YOLO network pruning method, which can be used as a lightweight mango detection model for mobile devices; and Cai et al. (2020) proposed an improved YOLO-V3 based on MobileNetv1 as a backbone to detect fish. The above research results show that good accuracy has been achieved in the detection of normal size targets with obvious differences from the background. However, for underwater feed pellets, the target is small and has high density, and the original YOLO algorithm still has limitations for small-target detection.

Therefore, this paper proposes an uneaten feed pellet detection method for aquaculture to address the problems of the low contrast, small size and large number of underwater feed pellets in images. Based on an improved YOLO-V4 (Bochkovskiy et al., 2020) network, the optimization design was carried out to improve the detection accuracy of uneaten feed pellets, thus providing useful input information for formulating a scientific pellet feeding strategy. The next sections of this paper are arranged as follows: Section 2 introduces the dataset, the related work on the improved algorithm and the improved network; Section 3 presents the results and discussion; and Section 4 is the conclusion.

2. Materials and methods

2.1. Dataset

2.1.1. Data acquisition and image features

The experimental data were collected from the 'Deep Blue No. 1' far-offshore mariculture net cage, which is located in the cold-water mass area of the Yellow Sea of China. Adult Atlantic salmon were raised in cages, and videos of the uneaten feed pellets were collected. The cage dimensions and image acquisition schematic are shown in Fig. 1. The cage had a circumference of 180 m and a height of 38 m. To better reflect the feeding status of fish, the area under the outlet of the feeder was photographed. To avoid the impact of vertical light, the camera shot from the bottom up at an angle of 60 degrees with the water surface. Considering changes in light, turbidity, etc., the camera used a zoom lens and was turned in automatic mode. The viewing angle of the lens was 2α, which was varied between 34 and 84 degrees; the typical maximum vision distance d was approximately 10 m; and the

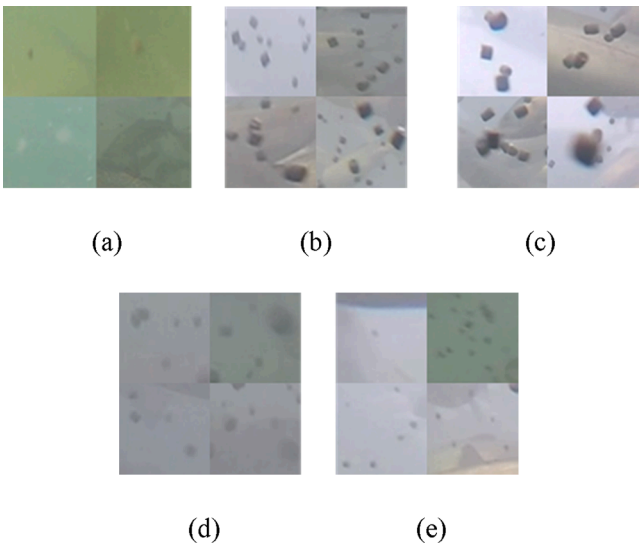


Fig. 2. Example pictures showing detection difficulties in local areas. (a) Turbid water; (b) high-density targets; (c) occlusion and adhesion; (d) pixel blur; (e) minuscule targets.

corresponding inspected area was the feeding area. The captured video size was 1920 × 1080, and the frame rate was 60 frames per second (FPS). The images with uneaten feed pellets were extracted from the video data as the dataset. The specific operation was to extract one frame every five seconds from the video. Considering that there were a large number of uneaten feed pellets in each picture, 40 images of uneaten pellets were finally selected.

In this paper, partial pictures (Fig. 2) were captured from the training dataset. There were great challenges in the task of detecting uneaten feed pellets, as follows:

(1) Turbid water: The images were blurred by fish excrement and floating objects in the water, as shown in Fig. 2(a);
(2) High density: A large number of small uneaten feed pellets increased the difficulty of detection and challenged the detection limit of the YOLO series algorithm, as shown in Fig. 2(b);
(3) Occlusion and adhesion: Occlusion and adhesion existed between feed pellets, as shown in Fig. 2(c), which could easily lead to missed detections;
(4) Pixel blur: Falling feed pellets were blurred targets, and in underwater imaging, the target pixels could be blurred by light, motion blur, the camera failing to focus, and so on (Fig. 2(d));
(5) Minuscule target: Feed pellets sufficiently far from the camera were extremely small targets, as shown in Fig. 2(e).

2.1.2. Image and data enhancement

To obtain a better recognition effect, the training images were preprocessed by contrast-limited adaptive histogram equalization (CLAHE) (Reza, 2004). CLAHE is an improvement of adaptive histogram equalization (AHE). AHE improves the image contrast by adjusting the brightness distribution, which makes it easy to amplify the noise in the image. On the basis of AHE, CLAHE appropriately limits the amplification of contrast to limit the interference caused by noise amplification. In this paper, the pretraining dataset was used for transfer learning as a data enhancement method. The mosaic in YOLO-V4 was also used for data enhancement. Mosaic is an extension of Cutmix (Yun et al., 2019). Cutmix is a mixture of two images, and Mosaic is a mixture of four images. This increases the batch size in the training process and correspondingly reduces the batches in the training process, thus reducing the hardware requirements.

2.1.3. Image annotation and dataset production

The open-source script *LabelImg* (https://github.com/tzutalin/labelImg) on GitHub was used to annotate the dataset. After running the *LabelImg* script, the target samples in each image were marked, and then, an XML file was generated containing the target type and coordinate information for training on the dataset. Because there were many targets in the uneaten feed pellet dataset, it was difficult to mark them. In addition, the complexity of the dataset caused difficulties for follow-up model training. A pretrained dataset consisting of underwater feed pellet pictures taken by mobile phones was prepared for transfer learning.

In the pretraining dataset, 175 pictures of underwater feed pellets were taken, with 7684 annotated samples. The improved network was used to train the pretraining dataset, and the parameter settings during training were as follows. The momentum was set to 0.949, the initial learning rate to 0.001, and the weight decay to 0.0005. The batch size was set to 16, and 6000 iterations were performed to generate the pretraining weight file. Forty underwater feed pellet pictures of far-offshore mariculture cages were used in the production of the dataset, and the annotated sample size was 7200. Among them, 35 pictures with 6007 annotated pellet samples were used as the training dataset, and 5 pictures with 1193 annotated pellet samples were used as the verification

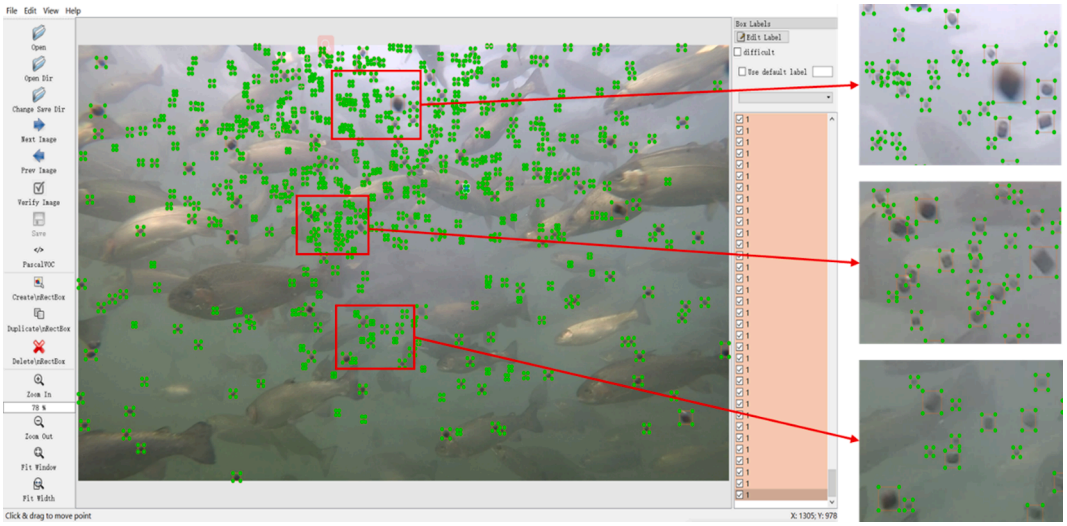


Fig. 3. Examples of annotated images.

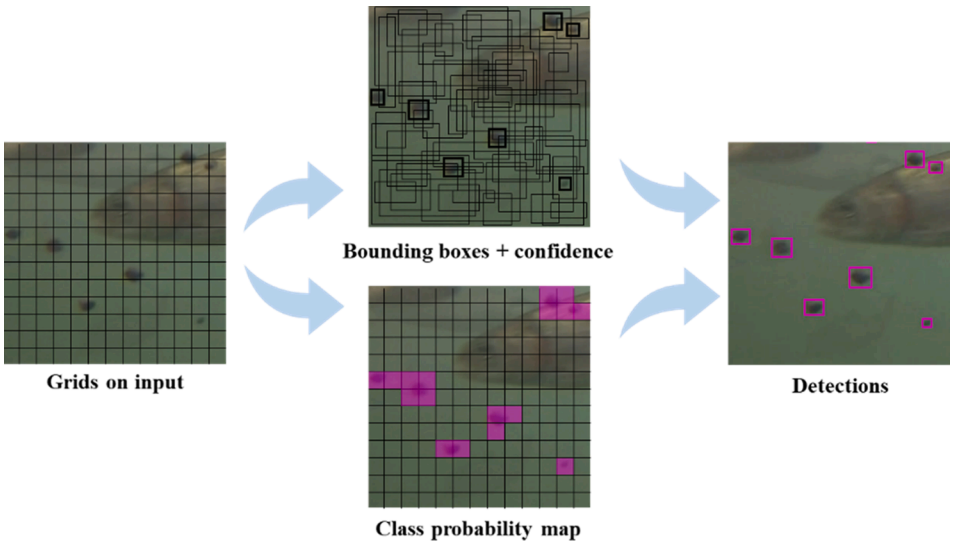


Fig. 4. The detection principle of the YOLO algorithm.

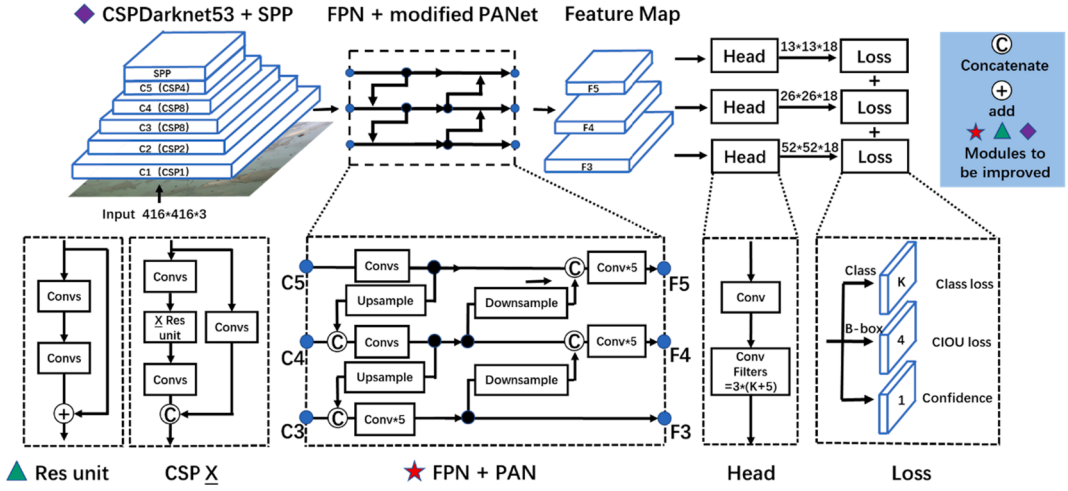


Fig. 5. Algorithm structure diagram of YOLO-V4.

dataset. During training, the weight file generated by the pretraining dataset was used to transfer and learn the training dataset. An example of the annotated images is shown in Fig. 3.

2.2. Related works

2.2.1. YOLO-v4

As a current excellent one-stage target detection algorithm, the YOLO-V4 network represents continued improvements to YOLO-V1 (Redmon et al., 2016), YOLO-V2 (Redmon and Farhadi, 2017), and YOLO-V3 (Redmon and Farhadi, 2018). The detection principle of all YOLO series algorithms is the same. Compared with the two-stage algorithm represented by Faster R-CNN (Ren et al., 2015), the region proposal network has been removed, which greatly improves the detection speed. As a classical one-stage algorithm, the detection problem is transformed into a regression problem. After the image is divided into S*S grids, the grids are used to predict whether there is a target in the grid, and the regression method is used to predict the location information and confidence. The basic principle is shown in Fig. 4. The confidence calculation formula is as follows:

$$Pr(Class_i|Object)*\Pr(Object)*IOU_{pred}^{truth} = \Pr(Class_i)*IOU_{pred}^{truth}, \Pr(Object) \in \{0,1\} \quad (1)$$

When the grid contains objects, Pr(Object) equals 1, and it is 0 otherwise. After the predicted box is generated, the final results are filtered by nonmaximum suppression.

The YOLO-V4 network is based on YOLO-V3 and has various degrees of optimization based on data processing, the backbone network, network training, the activation function, the loss function, and other aspects. YOLO-V4 uses the Mosaic data enhancement method CSPDarknet53 (Wang et al., 2020) as the backbone network and multiscale training as the network training method. Additionally, it uses feature pyramid pooling (He et al., 2015), the Mish activation function (Misra, 2019), and CIOU loss (Zheng et al., 2020) as the bounding box regression loss. The algorithm structure diagram of YOLO-V4 is shown in Fig. 5.

The loss function of YOLO-V4 consists of bounding box regression loss, confidence loss, and classification loss. *Loss(coord)* is bounding box regression loss. *Loss(conf)* is confidence loss. *Loss(cls)* is classification loss. The formula of the loss function is as follows:

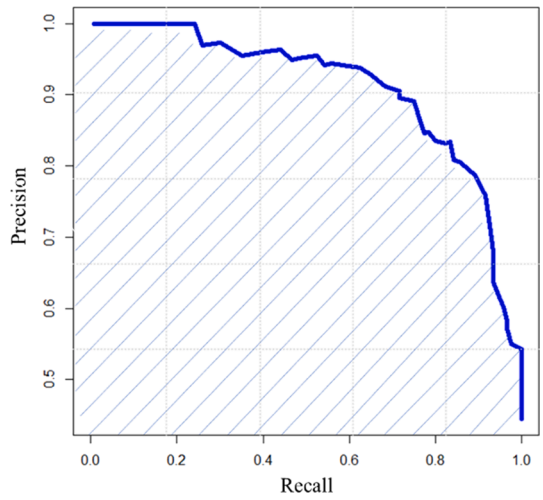


Fig. 6. Precision-recall curve.

$$Loss(object) = Loss(coord) + Loss(conf) + Loss(cls) \tag{2}$$

$$Loss(coord) = \lambda_{coord} \sum_{i=0}^{K*K} \sum_{j=0}^{M} I_{ij}^{obj} (2 - w_i * h_i)[L_{CIOU}] \tag{3}$$

$$Loss(conf) = -\sum_{i=0}^{K*K} \sum_{j=0}^{M} I_{ij}^{obj} \left[\widehat{C_i} log(C_i) + \left(1 - \widehat{C_i}\right) log(1 - C_i) \right] - \lambda_{noobj} \sum_{i=0}^{K*K} \times \sum_{j=0}^{M} I_{ij}^{noobj} \left[\widehat{C_i} log(C_i) + \left(1 - \widehat{C_i}\right) log(1 - C_i) \right] \tag{4}$$

$$Loss(conf) = -\sum_{i=0}^{K*K} I_{ij}^{obj} \sum_{c \in classes} [\widehat{p_i}(c) log(p_i(c)) + (1 - \widehat{p_i}(c)) log(1 - p_i(c))] \tag{5}$$

where $\lambda_{coord} = 5$, λ_{noobj}=0.5, I is the i-th cell of the feature map, k is the grid size, j is the prediction box responsible for the j-th box, and w and h represent the width and height of the ground truth. C_i is the confidence of the grid, I_{ij}^{obj} indicates whether there is an object in the i-th cell, and L_{CIOU} is the bounding box regression loss function.

Through principal analysis, the speed of the YOLO-V4 algorithm can largely meet the real-time requirements of the uneaten feed pellet detection task, but due to the small targets and the large number of them, the detection accuracy is not ideal. Therefore, by improving the YOLO-V4 network, this paper improves the detection accuracy and makes it suitable for underwater feed pellet detection tasks.

For target detection algorithms, performance evaluation indexes are needed to evaluate the algorithm model. According to the evaluation indexes of the neural network model, this paper uses the precision, recall, F1-score, and average precision (AP) as evaluation indexes (Goutte and Gaussier, 2005). The calculations of the precision, recall, and F1-score are shown in equation 6, 7 and 8. In the formulas, TP is the number of true positives, the samples that are correctly identified as feed pellets; FN is the number of false negatives, the samples that are incorrectly identified as background; TN is the number of true negatives, the samples that are correctly identified as the background; and FP is the number of false positives, the samples that are incorrectly identified as feed pellets.

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

$$F1 - score = \frac{2*Precision*Recall}{Precision + Recall} \tag{8}$$

The calculation formula for AP is shown as equation 9; r represents the integral variable, which is used to determine the integral of precision*recall and is between 0 and 1; AP refers to the area under the precision-recall (PR) curve, as shown in Fig. 6; AP_{50} is the average value of precision under different recall values when the intersection over union (IOU) is 0.5; and AP_{75} is the average value of precision under different recall values when the IOU is 0.75. $AP_{50:95}$ is the average of the ten values AP_{50}, AP_{55}… AP_{90}, AP_{95}.

$$AP = \int_0^1 Precision*Recalldr \tag{9}$$

$$AP_{50:95} = \frac{1}{10}(AP_{50} + AP_{55} + \cdots + AP_{90} + AP_{95}) \tag{10}$$

2.2.2. Panet

Feature pyramid networks (FPNs) (Lin et al., 2017) improve the detection accuracy by integrating the high- and low-layer features, which especially improves the detection accuracy of small targets. They continuously upsample the last layer of feature maps, combine them with the feature maps of each pyramid level, obtain new feature maps of different pyramid levels with a stronger representational ability, and then predict the category and location of targets in the feature maps. The path aggregation network (PANet) (Liu et al., 2018) improves the FPN. Considering the importance of the network's shallow feature information, it introduces the structure of bottom-up path augmentation so that the network can retain more shallow features.

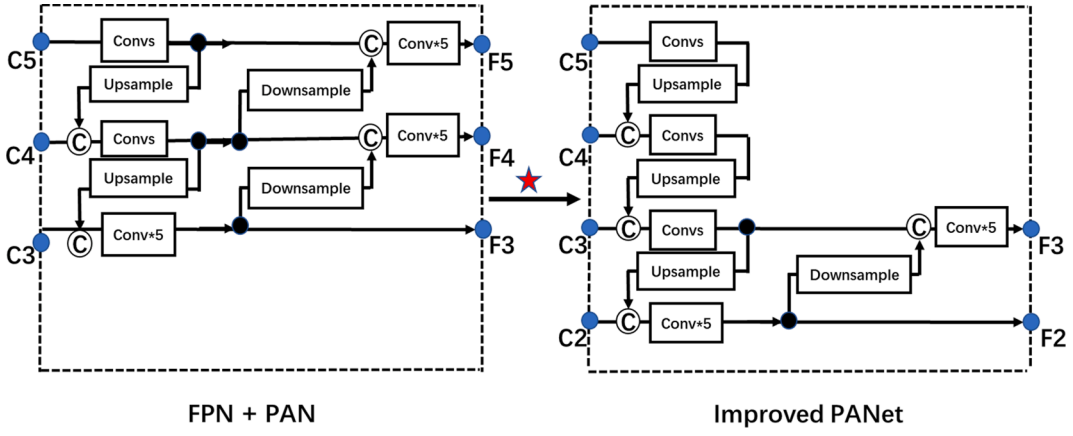


Fig. 7. Improvement of the feature map.

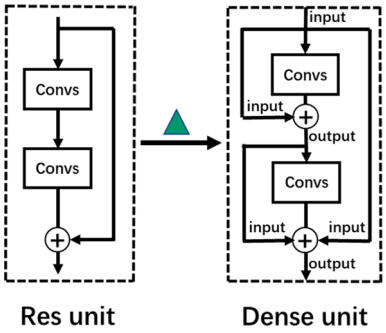


Fig. 8. Improvement of the residual blocks.

2.2.3. DenseNet

In contrast to ResNet (He et al., 2016), DenseNet (Huang et al., 2017) uses a densely connected mechanism in which all layers are interconnected. In the ResNet network, a specific layer is shortcut connected to a previous layer, whereas in the DenseNet network, each layer is connected with all the previous layers in the channel dimension and serves as the input to the next layer. Compared to the connection of ResNet, this is a dense connection to achieve feature reuse and improve efficiency.

2.3. The proposed algorithm

The proposed algorithm is an improvement of YOLO-V4 based on the characteristics of the dataset:

To address the excessive number of small targets, the PANet network connection was modified to obtain feature maps with finer-grained information and the feature map responsible for detecting large targets was pruned (Fig. 7);
To increase the training speed, the residual network was changed to a densely connected network. In this way, the feature transfer and reuse of the network were strengthened, and the problem of gradient disappearance in the training process of the dataset was solved (Fig. 8);
To increase the calculation speed, the number of YOLO-V4 network layers was reduced, and the residual blocks CPS1, CPS2, CPS8, CPS8, and CPS4 in CSPDarknet53 were changed to the dense connection blocks D-CPS1, D-CPS2, D-CPS 4, D-CPS4, and D-CPS2 to reduce the redundant features (Fig. 9).

2.3.1. Improvement of the feature map extraction network

In this study, the feature map of YOLO-V4 was improved by changing the PANet network structure. In addition, the network could retain more shallow features by fusing more shallow feature map information. Moreover, feature maps that were more conducive to small-target detection were generated to obtain richer fine-grained information. At the same time, the feature maps responsible for large-target detection were abandoned because of the small number of large targets in the dataset. The structure of the improved algorithm is shown in Fig. 7. The convolutional layer output of the backbone was upsampled an additional time, and its output was fused with the corresponding layer C2 in CSPDarknet53 to generate the F2 feature map. Moreover, the two downsampling operations at the end of the network were reduced, the F4 and F5 feature maps were removed, and the model calculation cost was minimized under the condition of ensuring the detection task.

2.3.2. Improvement of the residual blocks

For the residual module in CSPDarknet53 in YOLO-V4, the connection mode was modified to be dense. In the Res unit in YOLO-V4, the densely connected mechanism of DenseNet was introduced, and two shortcut connections were added. In this article, the new connection block is named the Dense unit. For small-target detection tasks, the Dense unit solves the problem of gradient disappearance and strengthens the feature transfer and reuse of small-target detection. The connection mode is shown in Fig. 8.

2.3.3. Improvement of the de-redundancy

Due to the large number of layers in the YOLO-V4 network, it is suitable for training difficult and complex datasets. For the underwater uneaten feed pellet dataset, there is only one class for detection. The focus is to solve the problem of small-target detection and feature reuse. Therefore, YOLO-V4 is too redundant for underwater feed pellet detection tasks. In response to this problem, this article modified the algorithm by performing de-redundancy operations on the backbone, which reduced the number of network layers and modified the convolution blocks CSP1, CSP2, CSP8, CSP8, and CSP4 in CSPDarknet53 to dense connection blocks D-CSP1, D-CSP2, D- CSP4, D-CSP4, and D-CSP2. This change reduced the number of redundant features and improved the calculation speed. The structure of this improvement is shown in Fig. 9.

3. Results and discussion

In this study, the Darknet framework was used to improve the YOLO-V4 network. The experimental environment is shown in Table 1.

Table 1
Experimental environment.

Configuration	Parameter
CPU	Intel Core i7-9700 K
GPU	Nvidia GeForce RTX 2080 Ti
Operating system	Windows 10
Accelerated environment	CUDA10.2 CUDNN7.6.5
Development environment	Visual Studio 2019
Library	Opencv3.4.0

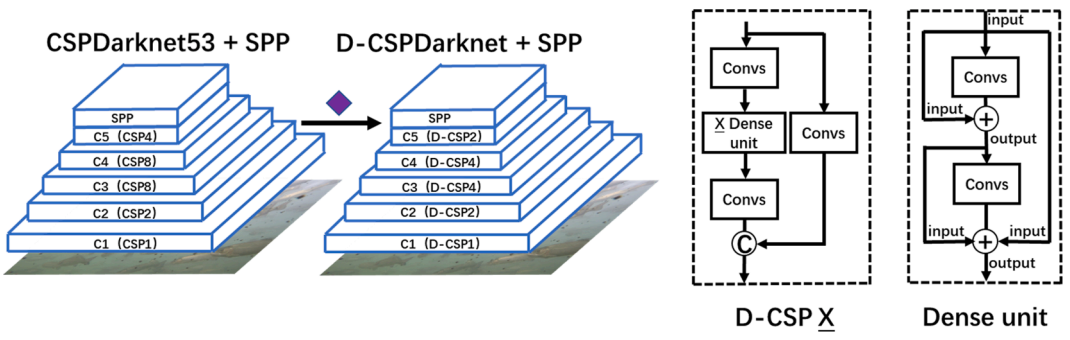


Fig. 9. Improvement of the de-redundancy.

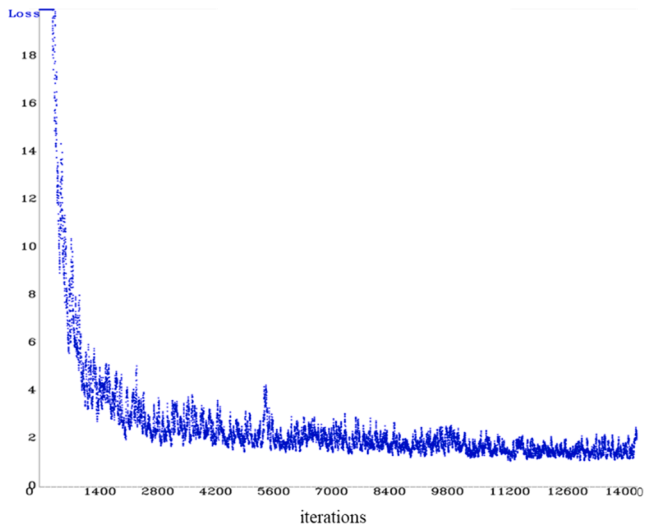


Fig. 10. Loss curve during training.

Table 2
Validation set experiment result; conf-thresh = 0.25; IOU = 0.5.

Conf-thresh = 0.25 IOU = 0.5	Precision	Recall	F1-score	TP	FP	FN
YOLO-v4	0.86	0.69	0.77	827	139	366
Improved YOLO-v4	0.94	0.89	0.91	1062	70	131

Considering the GPU memory limit during training, the batch size was set to 16. To better analyze the training process, 14,000 iterations were performed. The momentum, initial learning rate, weight decay, and other parameters refer to the original parameters in the YOLO-V4 network. The model was trained via the above settings; the learning rate dropped to 0.0001 after 10,000 steps and to 0.00001 after 12,000 steps.

In this paper, the trained improved YOLO-V4 model was tested to verify the performance of the algorithm. An image with the same resolution as the training input was tested. Since YOLO-V4 has five down samplings in the early stage, the size of the input image should be a multiple of 32. To verify the influence of the improved algorithm on the detection results of small targets, a more challenging resolution of 416*416 was selected for the experiment.

3.1. Training result analysis

The loss curve of the training process is shown in Fig. 10. As seen from the figure, before 1400 iterations, the loss decreased rapidly, and then with the increase in the iterations, the loss slowly oscillated downward and tended to be stable. The final loss value was stable at approximately 2. However, due to the complexity of the dataset itself, the loss value could not continue to decrease. Because of the gradual decrease and stabilization of the loss value, the model gradually converged, and the training results met expectations.

In this paper, the test results of the verification set were analyzed statistically, and the number of positive samples in the validation set was 1193. The conf-thresh (target confidence threshold) was taken as 0.25, and the IOU was taken as 0.5 for verification. The results are shown in Table 2. The precision, recall, and F1-score of the improved model improved to varying degrees. Specifically, the precision increased by 8%, the recall increased by 20% and the F1-score increased by 14%. The significant increase in TP and the significant decrease in FN were the main reasons for the increase in recall, which meant that more true positive samples had been correctly identified, indicating the effectiveness of the algorithm improvement.

3.2. Algorithm performance evaluation

3.2.1. Pretraining

A pretraining dataset consisting of 175 underwater feed pellet images was produced, and a pretraining weight file was generated for training. The local optimal solution of the dataset was obtained, and then the dataset was trained by transfer learning. Fig. 11(a) compares the results of using and not using the transfer learning strategy. The experiment showed that AP was increased by 3% with transfer learning, indicating that this strategy was effective.

3.2.2. Performance comparison of feature map extraction network

When using the original YOLO-V4 algorithm to detect uneaten feed pellets, frequent issues include false detection, missed detection, and inaccurate boundary box prediction. The areas with detected errors are mostly due to high density, overlaps, small targets, etc., which lead to a high false detection rate and high missed detection rate. The fundamental reason for the errors is that there is no feature output layer that can be responsible for small targets. The feature map of YOLO-V4 itself cannot be responsible for predicting the high density of small targets, leading to detection errors.

Aiming at detecting the small targets in the image, this study improved the feature map of YOLO-V4. A feature map that is more conducive to small-target detection was added, while the feature map that is responsible for large-target detection was discarded to reduce the amount of model calculation. The improved results are shown in Fig. 11 (b). After the feature map was improved, the AP_{50} of the verification set increased from 68.5% to 90.3% or 21.8%. The results showed that the improvements to the feature map for small targets greatly increased the detection accuracy. This was because the grid cells in the finer-grained feature layer were responsible for a smaller pixel area. Therefore, the finer-grained feature layer correctly detected the previously missed small and fuzzy targets while effectively improving the prediction accuracy of the bounding box and reducing false detections in the case of

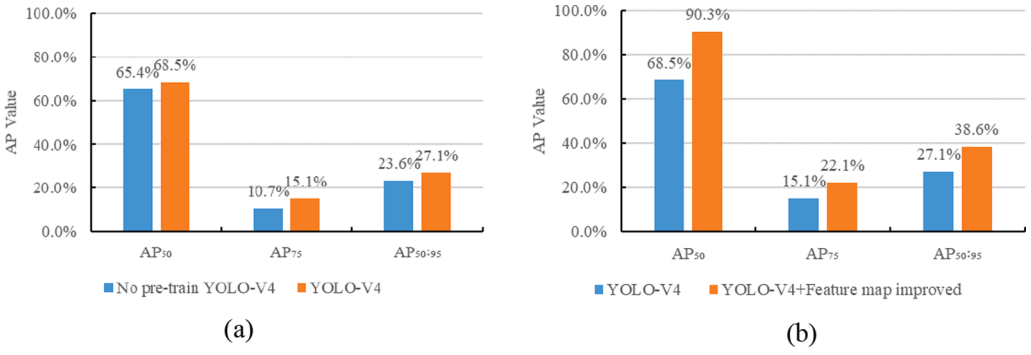


Fig. 11. Contrast graphs of the improved algorithm results. (a) Contrast graph of pretraining; (b) contrast graph of the feature map improvement.

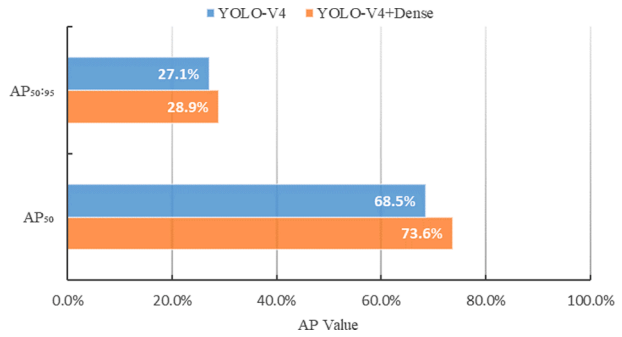
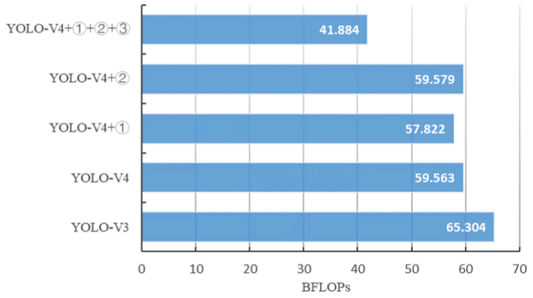


Fig. 12. Contrast graph of the dense connection improvement.



① Feature map improvement ② Dense connection ③ De-redundancy

Fig. 13. Contrast graph of de-redundancy.

Table 3
Contrast table of the algorithm results.

Algorithms	AP_{50}	AP_{75}	$AP_{50:95}$
YOLO-V3	46.98%	5.02%	15.08%
No pre-train YOLO-V4	65.40%	10.68%	23.57%
YOLO-V4	68.49%	15.12%	27.14%
YOLO-V4+①	90.30%	22.09%	38.55%
YOLO-V4+②	73.63%	13.15%	28.92%
YOLO-V4+①+②+③(Improved YOLO-V4)	**92.61%**	**22.34%**	**38.88%**

① Feature map improvement ② Dense connection ③ De-redundancy

occlusion.

3.2.3. *Performance comparison of improved dense connection unit*

The results of the dense connection improvements are shown in Fig. 12. After the connection was improved, the AP_{50} of the verification set increased from 68.5% to 73.6% or 5.1%. The purpose of dense connections is to decrease the gradient disappearance during training by increasing feature reuse. The modification of the residual blocks in the backbone of the YOLO-V4 network was based on the dense connection mode, and the added operation was the accumulation operation of the convolution kernel. The added calculation cost was very small and can be ignored, but it effectively improved the detection accuracy.

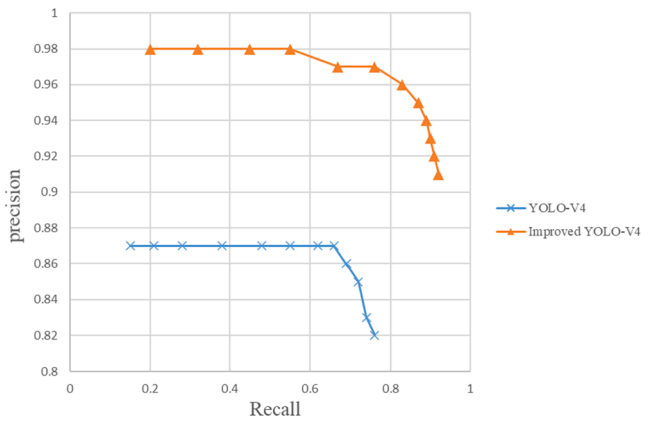


Fig. 15. PR curves of the models of pellets.

3.2.4. *Performance comparison of de-redundancy*

Feed pellet detection is a single-class target detection task, and the network structure of YOLO-V4 is too large and superfluous for the task. On the premise of ensuring the detection accuracy, the network can be appropriately pruned to improve the detection speed. The total computation for each algorithm in billion floating point operations (BFLOPs) is shown in Fig. 13. The results show that after the de-redundancy operation and feature map modification, the improved algorithm reduced the computation amount by approximately 30%. The de-redundancy operation reduced the number of network convolutional layers, so the total computation of the model was reduced accordingly.

3.3. *Performance comparison of the overall algorithm*

The results of different improvement strategies are compared in Table 3 and Fig. 14. Compared with YOLO-V3, the original YOLO-V4 has better target detection performance. After the ① feature map improvement, ② dense connection improvement, and ③ de-redundancy operation are performed on YOLO-V4, the detection accuracy for

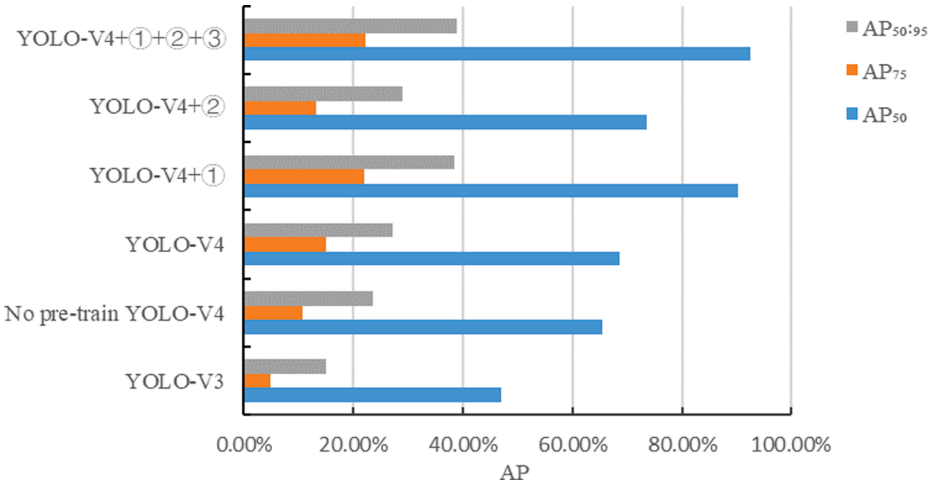


Fig. 14. Contrast graph of the algorithm results.

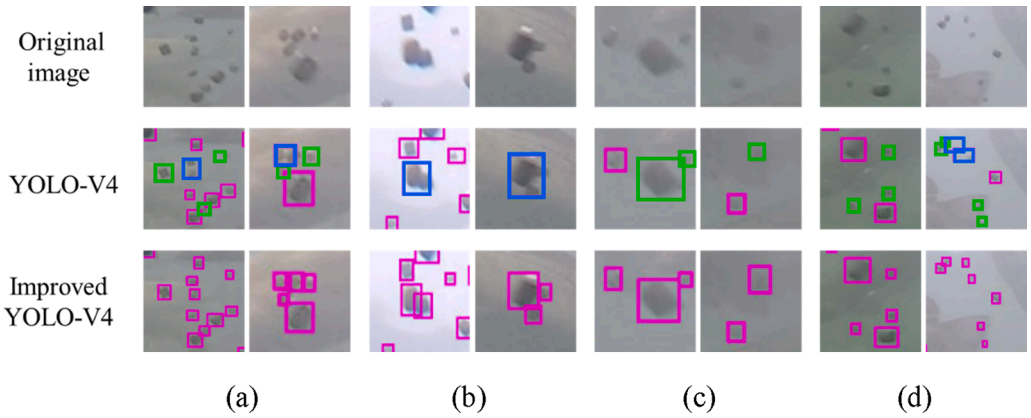


Fig. 16. Contrast diagram of partial test picture results. (a) High density; (b) occlusion and adhesion; (c) pixel blur; (d) minuscule target.

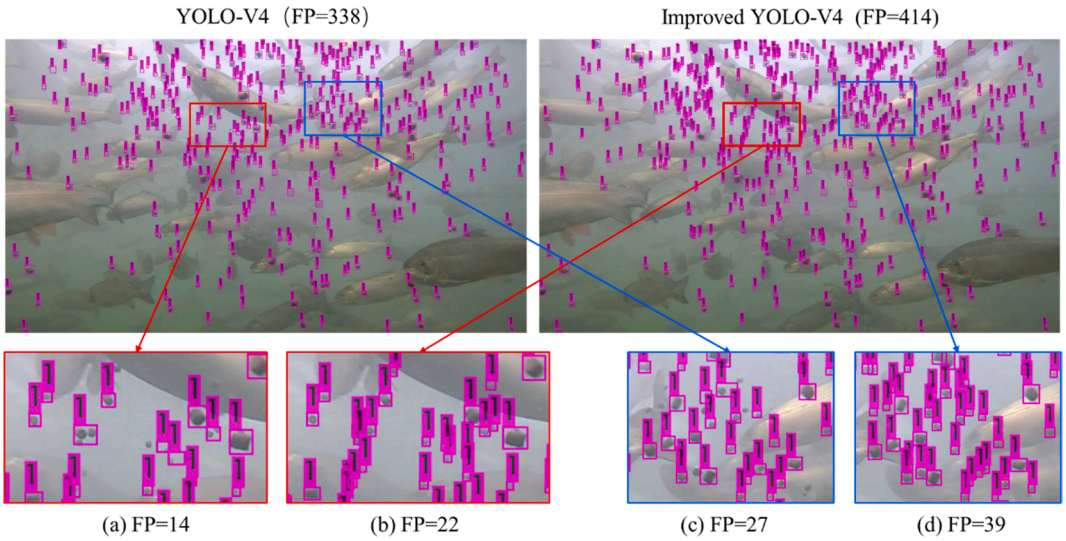


Fig. 17. Contrast diagram of the test picture results.

uneaten feed pellets is greatly improved. The results show that AP_{50} increased from 65.40% to 92.61%, with a difference of 27.21%. Through a series of improved operations, the detection accuracy for small targets is improved effectively.

Fig. 15 compares the pellet detection PR curves before and after the improvements to YOLO-V4. The enclosed area of the improved PR curve is larger than the area before the improvement. This means that the proposed algorithm achieved higher average accuracy.

Fig. 16 shows example images of the comparison results before and after improvement. The red box represents a correct detection, the blue box represents a false detection, and the green box represents a missed detection. Fig. 16(a) represents the detection under high density; Fig. 16 (b) represents the detection under the conditions of adhesion and occlusion; Fig. 16(c) is the detection in the case of pixel blur; and Fig. 16(d) represents the detection of minuscule targets. The results show that the improved YOLO-V4 network can effectively improve the detection accuracy for uneaten feed pellets. The reason for this is that the grid cell in the feature map extraction network of the original YOLO-V4 is responsible for the larger pixel area; in other words, a larger pixel area needs to predict more possible targets. Therefore, it prediction errors occur more easily in the case of high density, overlapping and small targets. The grid cell of the improved feature map extraction network can predict the target in a smaller pixel area, which effectively improves the detection accuracy. The improvement of the dense connection makes the features of the target in the fuzzy area richer through feature reuse, which is conducive to prediction.

Additionally, the test results of the whole picture before and after the improvement are compared in Fig. 17. Under the interference of factors such as high density and extremely small targets, the grid cells in the feature map extraction network are not accurate enough to predict the targets, resulting in the original YOLO-V4 has many missed and false detections. After the improvement, it still has good detection results, and the effect is clearly better than that of the unimproved method. Fig. 17 (a) and Fig. 17(b) represent the detection results for the same area (red box) of the same picture before and after improvement; Fig. 17(c) and Fig. 17(d) also represent the detection results for the same area (blue box) before and after improvement. By comparing them, it can be clearly seen that the improved detection results are more accurate.

4. Conclusion

To overcome the challenges of low-quality underwater images and extremely small targets in feed pellet detection, this paper presents an improved YOLO-V4 network to detect uneaten feed pellets in underwater images. The YOLO-V4 network is improved by modification of the feature maps using dense connections and a de-redundancy operation. The experimental results showed that the proposed improved YOLO-V4 network is superior to the original YOLO-V4 network. The AP_{50} increased from 65.40% to 92.61%, with a total increase of 27.21%. The improved network also reduced the amount of computation by

approximately 30%. This shows that the use of an improved feature map extraction network and dense connection is effective for improving the detection accuracy, and the use of the de-redundancy method is effective for improving the detection speed. Therefore, the proposed algorithm can effectively detect underwater feed pellets and can be used in an actual aquaculture environment. Moreover, the detection of uneaten feed pellets is only the first step of scientific feeding. In the future, the use of uneaten feed pellets and other parameters to formulate more scientific feeding strategies will be a research hotspot.

CRediT authorship contribution statement

Xuelong Hu: Supervision, Project administration, Funding acquisition. **Yang Liu:** Methodology, Software, Validation, Formal analysis, Investigation, Writing - original draft, Writing - review & editing, Visualization. **Zhengxi Zhao:** Methodology, Formal analysis. **Jintao Liu:** Methodology, Formal analysis. **Xinting Yang:** Supervision, Project administration, Funding acquisition. **Chuanheng Sun:** Supervision, Project administration, Funding acquisition. **Shuhan Chen:** Supervision, Funding acquisition. **Bin Li:** Supervision, Project administration, Funding acquisition. **Chao Zhou:** Conceptualization, Methodology, Investigation, Resources, Writing - original draft, Writing - review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The research was supported by the National Key Technology R&D Program of China (2019YFD0901004), the Beijing Natural Science Foundation (6212007), the Youth Research Fund of Beijing Academy of Agricultural and Forestry Sciences (QNJJ202014), the Jiangsu Province 7th Projects for Summit Talents in Six Main Industries, the Electronic Information Industry (DZXX-149, No. 110), and the National Natural Science Foundation of China (61802336).

References

Atoum, Y., Srivastava, S., Liu, X., 2015. Automatic feeding control for dense aquaculture fish tanks. Ieee Signal Proc. Let. 22 (8), 1089–1093.

Bochkovskiy, A., Wang, C. Y., Liao, H. Y. M., 2020. YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv preprint arXiv:2004.10934.

Cai, K. W., Miao, X. Y., Wang, W., Pang, H. S., Liu, Y., Song, J. Y., 2020. A modified YOLOv3 model for fish detection based on MobileNetv1 as backbone. Aquacult. Eng., 102117.

Chalamaiah, M., Hemalatha, R., Jyothirmayi, T., 2012. Fish protein hydrolysates: proximate composition, amino acid composition, antioxidant activities and applications: a review. Food Chem. 135 (4), 3020–3038.

Chen, L., Yang, X.T., Sun, C.H., Wang, Y.Z., Xu, D.M., Zhou, C., 2020. Feed intake prediction model for group fish using the MEA-BP neural network in intensive aquaculture. Information Processing in Agriculture 7 (2), 261–271.

Chiang, J.Y., Chen, Y.C., 2011. Underwater image enhancement by wavelength compensation and dehazing. Ieee T. Image Process. 21 (4), 1756–1769.

De Verdal, H., Komen, H., Quillet, E., Chatain, B., Allal, F., Benzie, J.A., Vandeputte, M., 2018. Improving feed efficiency in fish using selective breeding: a review. Rev Aquacult 10 (4), 833–851.

Fernandes, A.F.A., Turra, E.M., de Alvarenga, É.R., Passafaro, T.L., Lopes, F.B., Alves, G. F.O., Singh, V., Rosa, G.J.M., 2020. Deep Learning image segmentation for extraction of fish body measurements and prediction of body weight and carcass traits in Nile tilapia. Comput. Electron. Agr. 170, 105274.

Føre, M., Alfredsen, J.A., Gronningsater, A., 2011. Development of two telemetry-based systems for monitoring the feeding behaviour of Atlantic salmon (Salmo salar L.) in aquaculture sea-cages. Comput. Electron. Agr. 76 (2), 240–251.

Føre, M., Alver, M., Alfredsen, J.A., Marafioti, G., Senneset, G., Birkevold, J., Willumsen, F.V., Lange, G., Espmark, Å., Terjesen, B.F., 2016. Modelling growth performance and feeding behaviour of Atlantic salmon (Salmo salar L.) in commercial-size aquaculture net pens: Model details and validation through full-scale experiments. Aquaculture 464, 268–278.

Foster, M., Petrell, R., Ito, M.R., Ward, R., 1995. Detection and counting of uneaten food pellets in a sea cage using image analysis. Aquacult. Eng. 14 (3), 251–269.

Goutte, C., Gaussier, E., 2005. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation.In: European conference on information retrieval, Santiago de Compostela, Spain, pp. 345-359.

Han, J., Honda, N., Asada, A., Shibata, K., 2009. Automated acoustic method for counting and sizing farmed fish during transfer using DIDSON. Fisheries Sci. 75 (6), 1359.

Harsij, M., Gholipour Kanani, H., Adineh, H., 2020. Effects of antioxidant supplementation (nano-selenium, vitamin C and E) on growth performance, blood biochemistry, immune status and body composition of rainbow trout (Oncorhynchus mykiss) under sub-lethal ammonia exposure. Aquaculture 521, 734942.

He, K.M., Zhang, X.Y., Ren, S.Q., Sun, J., 2015. Spatial pyramid pooling in deep convolutional networks for visual recognition. Ieee T. Pattern Anal. 37 (9), 1904–1916.

He, K. M., Zhang, X. Y., Ren, S. Q., Sun, J., 2016. Deep residual learning for image recognition.In: Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, America, pp. 770-778.

Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K. Q., 2017. Densely connected convolutional networks.In: Proceedings of the IEEE conference on computer vision and pattern recognition, Hawaii, America, pp. 4700-4708.

Juell, J., 1991. Hydroacoustic detection of food waste—a method to estimate maximum food intake of fish populations in sea cages. Aquacult. Eng. 10 (3), 207–217.

Labao, A.B., Naval, P.C., 2019. Cascaded deep network systems with linked ensemble components for underwater fish detection in the wild. Ecol Inform 52, 103–121.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444.

Li, C.Y., Guo, J.C., Cong, R.M., Pang, Y.W., Wang, B., 2016. Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior. Ieee T. Image Process. 25 (12), 5664–5677.

Li, D.W., Xu, L.H., Liu, H.Y., 2017. Detection of uneaten fish food pellets in underwater images for aquaculture. Aquacult. Eng. 78, 85–94.

Lin, T.Y., Dollár, P., Girshick, R., He, K.M., Hariharan, B., Belongie, S., 2017. Feature pyramid networks for object detection.In. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2117–2125.

Liu, S., Qi, L., Qin, H.F., Shi, J.P., Jia, J.Y., 2018. In: Path aggregation network for instance segmentation.In. Salt Lake City, America, pp. 8759–8768.

Liu, Z.Y., Li, X., Fan, L.Z., Lu, H.D., Liu, L., Liu, Y., 2014. Measuring feeding activity of fish in RAS using computer vision. Aquacult. Eng. 60, 20–27.

Llorens, S., Pérez-Arjona, I., Soliveres, E., Espinosa, V., 2017. Detection and target strength measurements of uneaten feed pellets with a single beam echosounder. Aquacult. Eng. 78, 216–220.

Måløy, H., Aamodt, A., Misimi, E., 2019. A spatio-temporal recurrent network for salmon feeding action recognition from underwater videos in aquaculture. Comput. Electron. Agr. 167, 105087.

Masser, M., 1992. Management of recreational fish ponds in Alabama. Circular ANR-Alabama Cooperative Extension Service, Auburn University (USA).

Misra, D., 2019. Mish: A self regularized non-monotonic neural activation function. arXiv preprint arXiv:1908.08681.

Rauf, H.T., Lali, M.I.U., Zahoor, S., Shah, S.Z.H., Rehman, A.U., Bukhari, S.A.C., 2019. Visual features based automated identification of fish species using deep convolutional neural networks. Comput. Electron. Agr. 167, 105075.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection.In: Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, America, pp. 779-788.

Redmon, J., Farhadi, A., 2017. YOLO9000: better, faster, stronger.In: Proceedings of the IEEE conference on computer vision and pattern recognition, Hawaii, America, pp. 7263-7271.

Redmon, J., Farhadi, A., 2018. Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.

Ren, S. Q., He, K. M., Girshick, R., Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks.In: Advances in neural information processing systems, Montreal, Canada, pp. 91-99.

Reza, A.M., 2004. Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement. Journal of VLSI signal processing systems for signal, image and video technology 38 (1), 35–44.

Rola, W.R., Hasan, M.R., 2007. Economics of aquaculture feeding practices: a synthesis of case studies undertaken in six Asian countries. FAO Fisheries Technical Paper 505, 1.

Shi, R., Li, T.X., Yamaguchi, Y., 2020. An attribution-based pruning method for real-time mango detection with YOLO network. Comput. Electron. Agr. 169, 105214.

Simmonds, J., MacLennan, D.N., 2008. Fisheries acoustics: theory and practice. John Wiley & Sons.

Sun, X., Zhao, Z., Zhang, S., Liu, J., Zhou, C., 2020. Image Super-resolution Reconstruction Using Generative Adversarial Networks Based on Widechannel Activation. Ieee Access 8, 33838–33854.

Terayama, K., Shin, K., Mizuno, K., Tsuda, K., 2019. Integration of sonar and optical camera images using deep neural network for fish monitoring. Aquacult. Eng. 86, 102000.

Tian, Y.N., Yang, G.D., Wang, Z., Wang, H., Li, E., Liang, Z.Z., 2019. Apple detection during different growth stages in orchards using the improved YOLO-V3 model. Comput. Electron. Agr. 157, 417–426.

Wang, C. Y., Mark Liao, H. Y., Wu, Y. H., Chen, P. Y., Hsieh, J. W., Yeh, I. H., 2020. CSPNet: A new backbone that can enhance learning capability of cnn. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Virtual, pp. 390-391.

Wei, Y.G., Wei, Q., An, D., 2020. Intelligent monitoring and control technologies of open sea cage culture: A review. Comput. Electron. Agr. 169, 105119.
Wu, T.H., Huang, Y.I., Chen, J.M., 2015. Development of an adaptive neural-based fuzzy inference system for feeding decision-making assessment in silver perch (Bidyanus bidyanus) culture. Aquacult. Eng. 66, 41–51.
Yang, X.T., Zhang, S., Liu, J.T., Gao, Q.F., Dong, S.L., Zhou, C., 2021. Deep learning for smart fish farming: applications, opportunities and challenges. Rev Aquacult 13 (1), 66–90.
Yun, S., Han, D., Oh, S. J., Chun, S., Choe, J., Yoo, Y., 2019. Cutmix: Regularization strategy to train strong classifiers with localizable features.In: Proceedings of the IEEE International Conference on Computer Vision, Seoul, South Korea, pp. 6023-6032.
Zhang, S., Yang, X.T., Wang, Y.Z., Zhao, Z.X., Liu, J.T., Liu, Y., Sun, C.H., Zhou, C., 2020. Automatic Fish Population Counting by Machine Vision and a Hybrid Deep Neural Network Model. Animals 10 (2), 364.
Zhang, W.X., Xia, S.L., Zhu, J., Miao, L.H., Ren, M.C., Lin, Y., Ge, X.P., Sun, S.M., 2019. Growth performance, physiological response and histology changes of juvenile blunt snout bream, Megalobrama amblycephala exposed to chronic ammonia. Aquaculture 506, 424–436.
Zheng, Z. H., Wang, P., Liu, W., Li, J. Z., Ye, R. G., Ren, D. W., 2020. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression.In: AAAI, New York, America, pp. 12993-13000.
Zhou, C., Xu, D.M., Lin, K., Sun, C.H., Yang, X.T., 2018. Intelligent feeding control methods in aquaculture with an emphasis on fish: a review. Rev Aquacult 10 (4), 975–993.
Zou, Z. X., Shi, Z. W., Guo, Y. H., Ye, J. P., 2019. Object Detection in 20 Years: A Survey. arXiv preprint arXiv:1905.05055.