Journal of Computers Vol. 34 No. 3, June 2023, pp. 49-61
doi: 10.53106/199115992023063403004

Pearl Detection Based on PearlNet

Qiang Yuan, Shuai-Shuai Liu, Bang-Yu Wang, Dang-Wei Han, Sai-Nan Du, Da-Xu Zhao*

College of Optical, Mechanical and Electrical Engineering, Zhejiang A&F University, Hangzhou, China
1789530106@qq.com,2622497731@qq.com,2326982993@qq.com,736810679@qq.com,
994827298@qq.com,daxu@zafu.edu.cn

Received 9 June 2022; Revised 5 September 2022; Accepted 28 October 2022

Abstract. In this paper, we propose an algorithm model PearlNet and the corresponding detection dataset for freshwater pearls detection, to increase the Degree of Automation and improve the efficiency of existing detection methods based on pearl colors and shapes. PearlNet based on CenterNet. According to the characteristics of the small target of freshwater pearls, the minimum size module of the network is deleted, and the attention mechanism is added at the same time, ignoring the irrelevant background information and focusing on the pearl feature information, which improves the accuracy of recognition. In the transport convolution process, the image quality effect caused by upsampling is reduced by data fusion. The experimental results proved that the PearlNet has a recognition accuracy of 98.4%, which is 15.43%, 9.05% and 5.2% higher than that of CenterNet, Yolo V3 and SSD. PearlNet can accurately identify the color and shape of pearls, which provides a reference for freshwater pearl identification and detection.

Keywords: pearl, PearlNet, CenterNet, object detection

1 Introduction

Freshwater pearls are deeply loved by consumers in the international market because they have gentle feelings different from other gemstones [1]. The cultivated freshwater pearls contain different colors and shapes that need to be sorted. Accurate identification of the color and shape of pearls can provide effective information for pearl sorting. Due to the current situation, the sorting of pearls mainly relies on manual sorting. The traditional manual sorting is costly, time-consuming, and inefficient, the accuracy of pear sorting is also easily affected by the level and status of inspectors, which limits the development of the freshwater pearl industry to a certain extent.

Due to the influence of geographical location and historical reasons, the cultivation of freshwater pearls is mostly in China, and China's research on freshwater pearls is relatively advanced. Researchers usually use optics or by analyzing different pearl image pixels to find the segmentation threshold, and furthermore adopt a manually selected threshold method for image segmentation to classify pearls. These methods all require high lighting conditions. Literature [2] by establishing the HSV model of pearls and extracted by Ostu Segmentation and the average value of H (Hue) and S (Saturation) according to the histogram of V (Value) weigh to distinguish the hierarchy. The method only distinguishes the pearl shape. Literature [3] proposes a method that assesses and quantifies each factor of the pearl quality by using optical measurement on pearl's body color, interference color, glossiness assesses and quantifies. The accuracy of this method depends on the lighting conditions and the cost is high, so it is not suitable for industrial application. In recent years, object detection algorithms based on deep learning have become very popular. Literature [4] uses the generative adversarial network to expand pearl dataset, and then use convolutional neural network to train and classify pearls still has the problem of low accuracy.

In recent years, CenterNet algorithm in object detection has emerged. CenterNet does not rely on complex decoding strategy or heavy head design, it can surpass the popular real-time detector, and has faster reasoning speed. Aiming at the task of freshwater pearls detection, PearlNet is proposed to improve CenterNet network, which fills the gap of freshwater pearls in target detection direction recognition [5]. The main contributions of this paper can be summarized as follows:

•We propose an algorithm model PearlNet for pearl detection, which is based on CenterNet. PearlNet has a recognition accuracy of 98.4%, which is 15.43%, 9.05% and 5.2% higher than that of CenterNet, Yolo V3 and SSD.

* Corresponding Author

•For the pearl detection task, this paper proposes a pearl-based object detection dataset, which fills the gap of object detection in the field of pearl detection.

•We propose a data fusion method to solve the problem of information in the deconvolution process and reduce the image quality impact caused by upsampling, The experimental results show that the method is effective.

2 Relative Work

In recent years, object detection based on deep learning has attracted more and more attention due to its wide application and technological breakthroughs in recent years. This task has been extensively studied for both academic and real-world applications, such as surveillance security, autonomous driving, traffic monitoring, UAV scene analysis, and robot vision [6]. Literature [7] proposed R-CNN is the pioneering work of deep learning for object detection that Using CNN to perform feature extraction on the region of interest, which solves the problem of insufficient expressive capability of features in object detection, but there still remains problems such as information loss and high computational cost. With the continuous development of deep learning and computer technology, many target algorithms based on deep learning such as FAST-RCNN [8], FASER-CNN [9], and YOLO [10] have been proposed. These algorithms are based on the Anchor Box algorithm, by creating a series of detection boxes (anchor boxes) with different sizes and aspect ratios, and then applying classification or regression on the content of the anchor boxes. The experimental results showed that anchor box has many shortcomings, including lack of generalization ability, high training cost, redundant calculation, and poor abnormal object detection ability. Aiming at the shortcomings of the anchor frame algorithm, anchor-free target detection algorithms have been proposed, for example CornerNet [11], CenterNet [12], etc. Literature [12] proposed the CenterNet (Object as Point) algorithm, which no longer needs to remove redundant borders. First, the algorithm adopts a keypoint-based idea to predict the center point of an object using a heatmap. Then, the idea of center prediction is adopted to predict the length and width of the bounding box, which is achieved by using the features obtained at the predicted center point. The application of this hybrid method obviates the need to use non-maximum suppression NMS for removing redundant structures, and as a result, the calculation is more concise than the existing target detection algorithm.

CenterNet is widely used in agriculture and industry. Literature [13-14] improved CenterNet to detect pedestrians by adding attention mechanism and normalization processing and achieved good results. Literature [15] proposed a vehicle detection method based on CenterNet model in deep learning. The experimental results show that has a good detection effect for vehicles in actual scenes, and the network has certain robustness. Literature [16] proposed the application of CenterNet in weed identification and combined with traditional vision, which got the best results.

In view of the current freshwater pearl detection problem, this paper introduces the target detection algorithm--CenterNet for the first time, with high robustness. Traditional visual methods are difficult to be effective in actual sorting due to environmental factors, such as shadows, lighting and other conditions. The deep learning method has the advantage of not relying on the environment, and only needs enough data sets to solve the cost problems such as lighting and shadows. The effectiveness of the proposed method is also verified by the work in this paper, and the experimental results show that the accuracy is 98.4%.

3 Methods

For the pearl detection task, we designed a target detection network PearlNet based on CenterNet. The network model is mainly composed of three parts as shown in Fig. 1. A backbone with ResNet50 network structure which is used to extract image features through down-sampling convolution. The middle feature processing stage which the feature is up-sampled to obtain the high-resolution feature image through cubic transposed convolution.

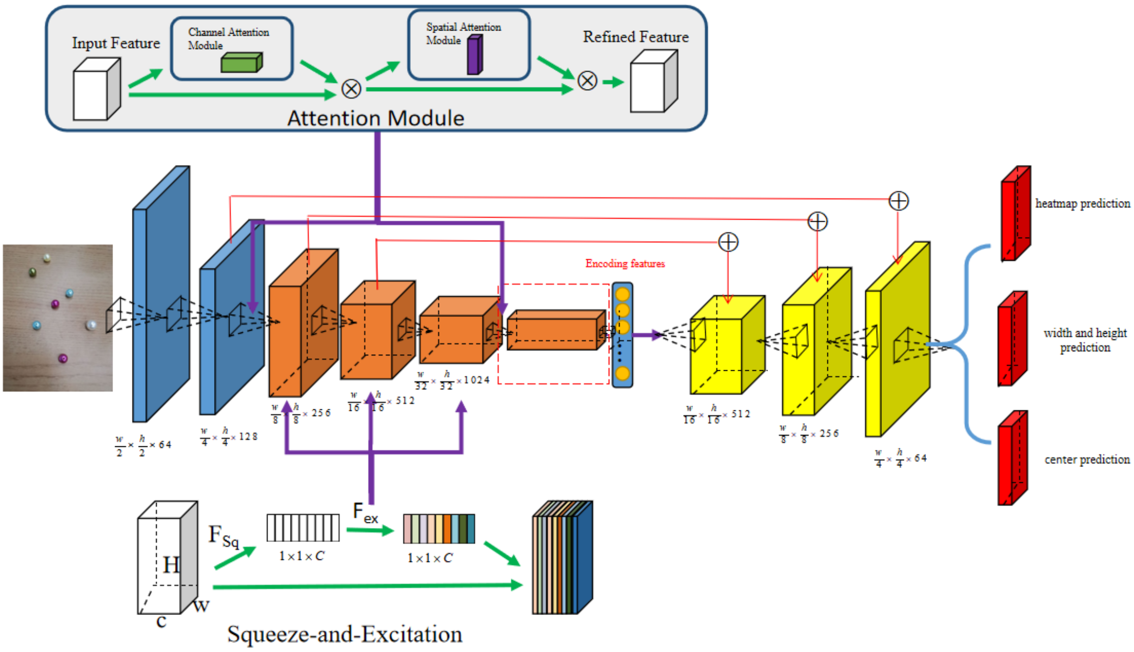


Fig. 1. PearlNet structure

3.1 Optimize the Backbone Feature Extraction Network

The backbone extraction network of PearlNet is based on ResNet50. The size of the Pearl dataset is concentrated at 720×960. We changed the network input from the original to and reduced it by the undistorted resize ratio. The features extracted from the picture are more informative. The initial backbone network is changed from one convolution to two convolutions to extract features.

The last layer of ResNet50 reduces the image to of the original, as shown in the red dotted box in (Fig. 1). The $^1\!/_{32}$ pixel of the original pearl image cannot contain the full information of a pearl, which impacts the pearl shape detection, thus affecting the final detection result. Therefore, we delete the Bottleneck network in the last layer of resnet50, which not only reduces the parameters of the feature extraction network model but also increases the integrity of the pearl shape feature.

3.2 SE (Squeeze-and-Excitation) Channel Attention Mechanism

The resnet50 optimized by PearlNet contains three bottleneck structures. The bottleneck structure is mainly composed of a convolution stacking and residual addition. Features are extracted by 3×3 convolution, and then the number of channels is adjusted by 1×1 convolution. In the process of convolution stacking, we focus on spatial information fusion and ignore the fusion between channels. By adding SE module to each bottleneck structure, the model can automatically learn the importance of different channel features. The SE module compresses the feature map of $h \times w \times c$ into $1 \times 1 \times c$, uses two fully connected layers to predict the importance of each channel, obtains the importance of different channels, and then weights the importance of each channel. operation, and finally multiply the learned activation value of each channel by the channel of the previous feature map.

3.3 CBAM Attention Mechanism

The CBAM (Convolutional Block Attention Module) model combines the channel and Spatial information of the feature map, without changing the size of the feature map, makes the model having a stronger perception ability of target information and suppressing the interference caused by invalid information. As shown in Fig. 2, the Channel Attention Module of CBAM uses maxpool and avgpool at the same time, goes through the MLP multilayer perceptron, uses the final results for two channels, and gets Channel attention consequences by using sigmoid function.

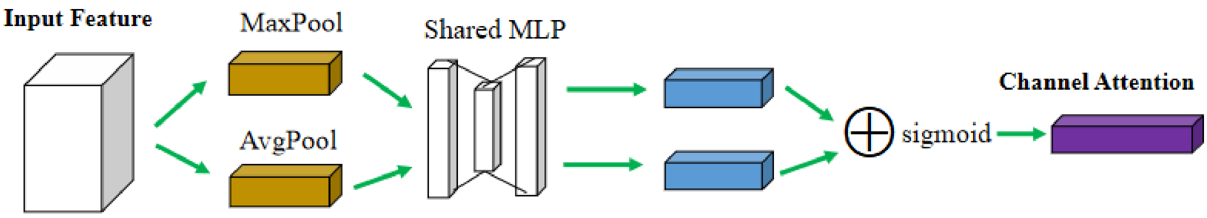


Fig. 2. Channel Attention Module in CBAM

As shown in Fig. 3, the Spatial Attention Module of CBAM reduces the dimension of the Channel Attention itself obtained by the channel, obtains the results of maxpool and avgpool respectively, splices the two results into a feature map, and then via convolution, the result after convolution by means sigmoid obtains the Spatial Attention result.

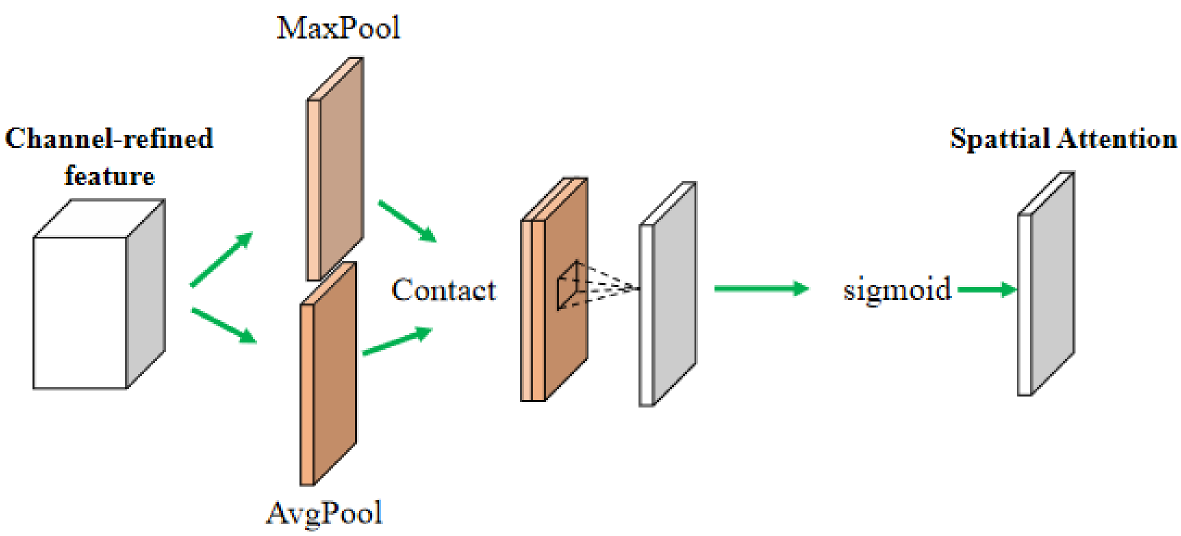


Fig. 3. Spatial Attention Module in CBAM

3.4 Optimize the Transposed Convolution Network Structure

Transposed convolution enlarges the original image so that it can be displayed at a higher resolution, but it will affect the image quality during the transposed convolution process, which will affect our test results. In order to reduce the impact of image quality caused by transposed convolution, we changed the number of channels of transposed convolution three times, so that the channel numbers are the same as the initial convolution layer and the first two bottleneck extraction feature networks, and then adding the features and previous features together for normalization.

3.5 Loss Function Setting

The loss function of PearlNet mainly refers to that of the CenterNet algorithm. PearlNet presents the target through the center point, and then regresses the target nature at the center point position so that the target detection problem transforms into a key point estimation problem. PearlNet scatters keypoints onto the heatmap through the Gaussian distribution function of (eq.1):

$$Y_{X,Y,C} = exp(-\frac{(x-\tilde{p}_x)^2 + (y-\tilde{p}_y)^2}{2\sigma_p^2}) . \quad \textbf{(1)}$$

Where

σ_P = standard equation

$\tilde{p}_x$, $\tilde{p}_y$ = coordinates of the corresponding key points in the image

The heat map of PearlNet is the original $1/4$, so there is a bias of center point. During the training process, the width and height of the prediction frame and the bias in the center point need to be regressed at the same time. Therefore, the total loss of the PearlNet model prediction includes the loss of the heat map (eq.2), the bias loss of the center point (eq.3), and the loss of width and height of the target frame (eq.4), which is defined as:

$$L_K = \frac{-1}{N}\sum \begin{cases} (1-\hat{Y}_{xyc})^{\alpha} \log(Y_{xyc}) & Y_{xyc}=1 \\ (1\text{-}Y_{xyc})^{\beta} (\hat{Y}_{\text{xyc}})^{\alpha} \ln(1\text{-}Y_{xyc}) & \text{otherwise} \end{cases} . \quad \textbf{(2)}$$

$$L_{size} = \frac{1}{N}\sum_{K=1}^{N}\left|\hat{S}_{PK} - S_K\right| . \quad \textbf{(3)}$$

$$L_{off} = \frac{1}{N}\sum_{P}\left|\hat{O}_P - (\frac{P}{R} - \tilde{p})\right| . \quad \textbf{(4)}$$

Where

N = number of key points

α , β = hyperparameter

$\hat{Y}_{\text{xyc}}$ = the Gaussian distribution function of (eq.1)

P = the true center point

$\hat{S}_{PK}$ = the network prediction width and height loss

R = the position of the figure subsampled

S_K = the width and height of the detection frame

The total loss of PearlNet is the weighted sum of the loss of the heat map, the bias loss of the center point and the loss of the width and height of the target box, which is defined as:

$$L_{ALL} = L_K + \lambda_{off} L_{off} + \lambda_{size} L_{size} . \tag{5}$$

Where

λ_{off} , λ_{size} = the scale factor

4 Results and Discussion

4.1 Evaluation Indicators

Pearl's test set will have four possible outcomes in Table 1 in PearlNet:

Table 1. The test set predicts

		Prediction	
		1	0
True	1	TP (True Positive)	FP (False Positive)
	0	FN (False Negative)	TN (True Negative)

Precision is calculated as the ratio of actual positive samples over the positive samples:

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

Recall predicts the ratio of the positive sample in the actual positive sample:

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

To measure the quality of a model, a precision or a recall alone is not sufficient, as they are mutually exclusive, usually with high precision corresponding to low recall, or low precision corresponding to high recall. Therefore, the AP is introduced to represent the area under the two curves of precision and recall, and mAP is the average of multiple categories of AP. Therefore, this paper chooses the mean precision as the measure.

4.2 Datasets

The current pearl recognition uses the traditional artificial threshold to identify the pearl color. The dataset obtained this way puts high demand on lighting conditions. During the production process, a closed photo studio is built for image collection, which substantially increases the workload, and the resulting model usually applies only to this batch of pearls, with poor robustness. When applied to the identification link in large-scale industrial sorting, this method is very costly, and the lighting conditions are undoubtedly a huge challenge. The image collection based on the PearlNet target detection algorithm used in this paper can be applied on both sunny and cloudy days under the indoor lighting and natural light collection scenario. As can be seen in Fig. 4, the collected images have shadows, etc. which are in line with the actual conditions.

Fig. 4. Pearl datasets

4.3 Training Process

The experimental hardware is a PC with 16GB memory and NVIDIA 2080ti GPU, the operating system is Ubuntu 18.04, and the algorithm model is programmed in python based on Pytorch deep learning framework. Some parameters are set as follows: the learning rate is $1*10^{-3}$, the momentum coefficient is 0.9, the weight decay coefficient is 0.0005, batch size is 8 and the number of iterations is set to 200. Fig. 5 shows the training loss functions of PearlNet and CenterNet. PearlNet is more suitable for freshwater pearl detection, and the loss is easier to converge. Compared with CenterNet, PearlNet has lower training loss, easier convergence, and is more suitable for pearl detection.

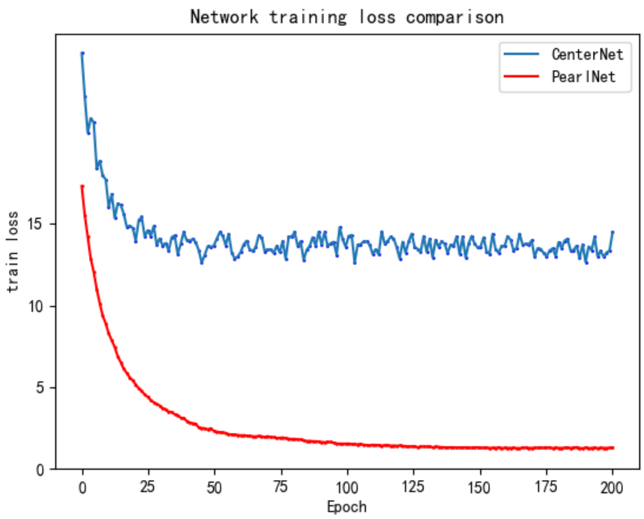


Fig. 5. Network training loss comparison

4.4 Experimental Results

The same dataset was used for training in the same batch on the CenterNet, Yolo V3 and SSD. The detection results are shown in Table 2, and compared with PearlNet. In terms of comprehensive index mAP, PearlNet's mAP is 98.40%, which is 15.43%,9.05% and 5.2% higher than CenterNet, Yolo V3 and SSD, respectively. Compared with CenterNet results, PearlNet's AP is much higher than CenterNet, in both color or shape recognition.

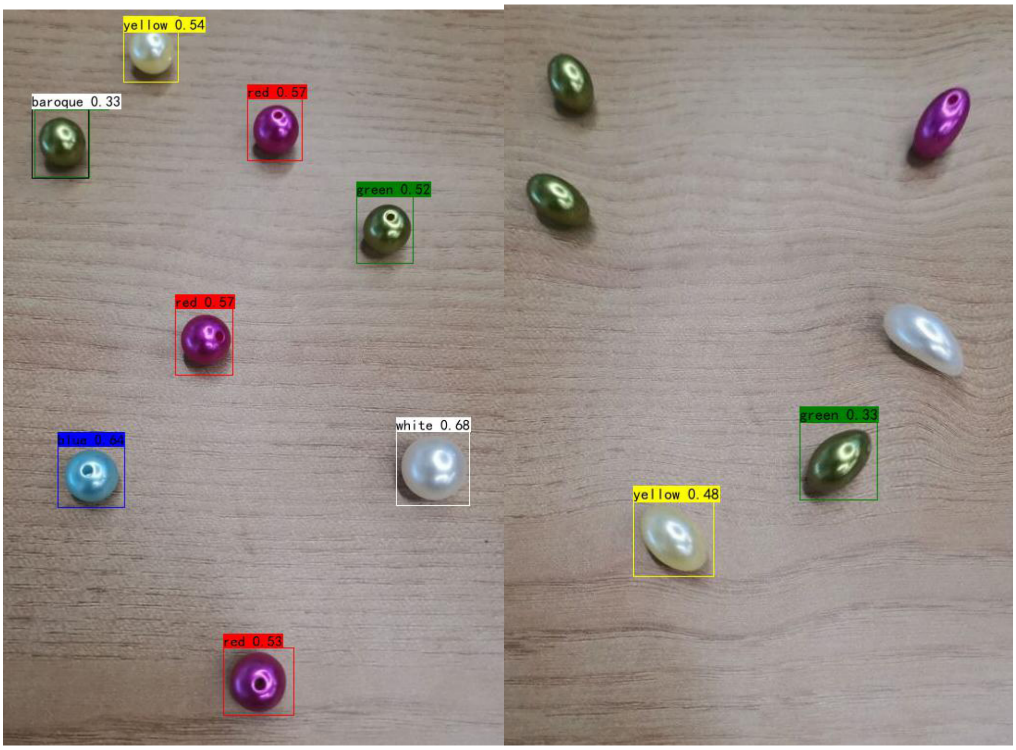
Table 2. Experimental results of pearl recognition with different object detection models

Label / Network	White AP/%	Red AP/%	Yellow AP/%	Green AP/%	Blue AP/%	Baroque AP/%	mAP/%	Time[a]
PearlNet	99.95%	99.75%	97.79%	97.79%	97.53%	99.77%	98.40%	32s
CenterNet	97.72%	77.25%	77.63%	71.40%	80.58%	93.20%	82.97%	20s
Yolo V3	99.99%	92.96%	99.15%	84.72%	90.64%	68.67%	82.97%	15s
SSD	99.26%	98.47%	98.58%	83.48%	89.92%	96.78%	93.92%	12s

[a] Time is the total detection time of the test set

Fig. 6 shows the detection results of the same image in different object detection models. As shown in Fig. 6(a), although YOLO V3 has a fast detection speed, the pearls with white and yellow labels in YOLO V3 are not recognized, and there also exists the problem of retaining multiple recognition boxes after individual non-maximal suppression. As shown in Fig. 6(b) for the CenterNet algorithm, the recognition ability of special-shaped pearls is poor. As shown in Fig. 6(c) is only good at shape recognition. As shown in Fig. 6(d), PearlNet is accurate in the prediction result, and the recognition of pearl shape and color is accurate. Compared with CenterNet and YOLO V3, the accuracy is further improved.

(a) Yolo V3



(b) CenterNet

(c) SSD

(d) PearlNet

Fig. 6. Pearls detection results for different models

4.5 Comparison of Ablation Results

To verify the effectiveness of the method proposed in this paper, “ablation studies” is used to compare the impact of different modules in the network on its performance.

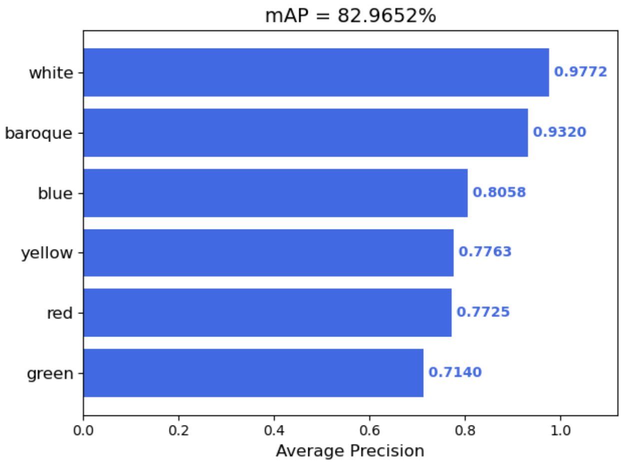
Table 3. Ablation results

Methods / Experiment	Optimize Backbone Network	SE Attention Mechanism	CBAM Attention Mechanism	Optimize Transposed Convolution	mAP/%	Time[a]
1	×	×	×	×	82.97%	20s
2	√	×	×	×	94.55%	23s
3	√	√	×	×	94.96%	24s
4	√	√	√	×	95.65%	32s
5	√	√	√	√	98.4%	32s

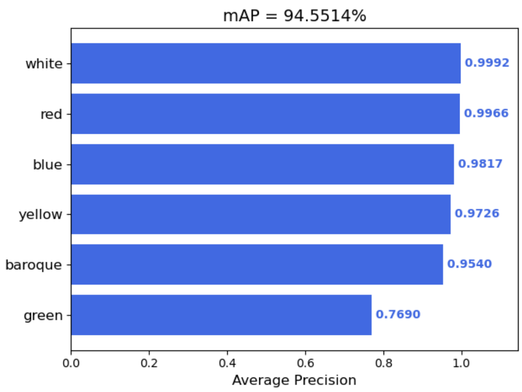
[a] Time is the total detection time of the test set

From the experimental results in Table 3, it can be seen that the influence of introducing different improvement strategies on the detection effect has different degrees of improvement. Experiment 1 and Experiment 5 are the experimental results of CenterNet and PearlNet, respectively. Experiment 2 only optimized the backbone feature extraction network, and the detection accuracy was also greatly improved. Since the input scale of the network is different, the detection time for the test set becomes more. Experiments 3 and 4 introduced different attention mechanisms, and the detection accuracy was improved to different degrees. Compared with experiment 5, experiment 4 only introduces data fusion in the transport convolution process, which improves the accuracy.

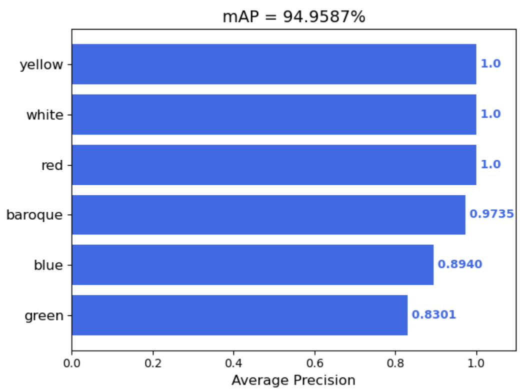
Fig. 7 shows the AP size of each category on the test set in the module ablation experiment. As can be seen from the Fig. 7(a) and Fig. 7(e), for freshwater pearl baroque, which AP of experiment 1 is only 71.40%, and the AP of Experiment 5 is 95.61%, compared with experiment 1, the AP detection accuracy of baroque has increased by 24.21%, and other categories have improved to varying degrees. PearlNet pays more attention to pearl shape and color features, which verifies the effectiveness of each module of the improved algorithm in this paper.



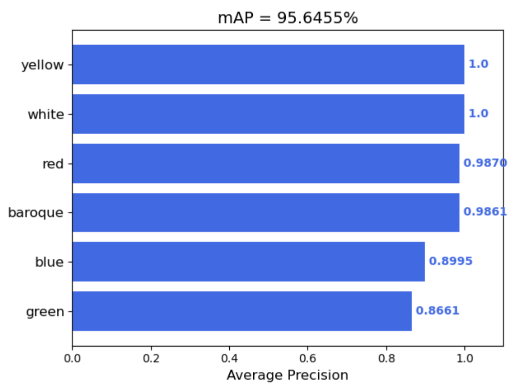
(a) mAP = 82.9652%



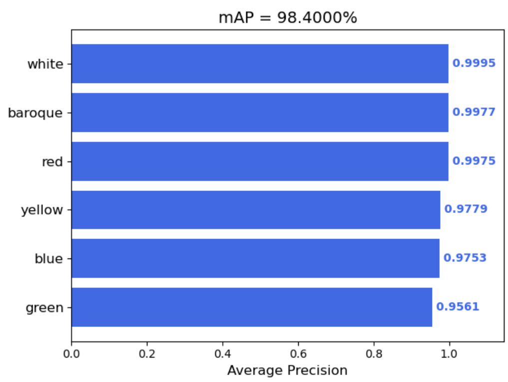
(b) mAP = 94.5514%



(c) mAP = 94.9587%



(d) mAP = 95.6455%



(e) mAP = 98.4000%

Fig. 7. Accuracy of each category of ablation experiments

5 Conclusion

In this paper, we propose an algorithm model PearlNet and the corresponding detection dataset for Freshwater pearl detection, to solve poor Degree of Automation and poor efficiency of the existing pearl color and shape detection. PearlNet based on CenterNet. According to the characteristics of the small target of freshwater pearls, the minimum size module of the network is deleted, and the attention mechanism is added at the same time, ignoring the irrelevant background information and focusing on the pearl feature information, which improves the accuracy of recognition. In the transport convolution process, the image quality effect caused by upsampling is reduced by data fusion. Through comparative experiments and ablation experiments, the effectiveness of the algorithm is verified.

In future work, the model will be further improved, through techniques such as channel pruning and parameter quantity. The number of parameters will be reduced and the detection speed will be improved, so that it can be more applicable in industry.

6 Acknowledgement

Fund Project: 1. Zhejiang Non-profit technology research program, sponsored by Natural Science Foundation of Zhejiang Province, LGN20E050003. 2. Provincial key research project, sponsored by Science Technology Department of Zhejiang Province, 2019C02003. We appreciate this project being supported by the above funds.

References

[1] G. Mondonneix, S. Chabrier, J.-M. Mari, A. Gabillon, Tahitian Pearls' Luster Assessment Automation, in: Proc. 2017 IEEE Applied Imagery Pattern Recognition Workshop, 2017.
[2] B. Li, G. Li, Y. Yang, X. Wang, A classification method of pearl shape based on Zernike moment, in: Proc. 2007 International Conference on Wavelet Analysis and Pattern Recognition, 2007.
[3] Y. Ota, H. Higashi, S. Nakauchi, Objective assessment and quantification of pearl quality by spectral-spatial features, in: Proc. 2015 2nd International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA), 2015.
[4] Q. Xuan, Z. Chen, Y. Liu, H. Huang, G. Bao, D. Zhang, Multiview Generative Adversarial Network and Its Application in Pearl Classification, IEEE Transactions on Industrial Electronics 66(10)(2019) 8244-8252.
[5] A. Raghunandan, Mohana, P. Raghav, H.V.R. Aradhya, Object detection algorithms for video surveillance applications, in: Proc. 2018 International Conference on Communication and Signal Processing (ICCSP), 2018.
[6] J.-B. Zhou, R.-Q. Zhang, K.-K. Hao, H.-B. Li, P.-W. Li, R.-H. Liu, X.-L. Guo, X. Zheng, Y.-C. Yang, H.-W. Xu, G.-K. Chen, K.-S. Dai, Z.-H. Jiao, X.-B. Cai, Two-Stage Quality Assessment Method of Power Equipment X-Ray Image Based on Improved CenterNet, in: Proc. 2021 IEEE Far East NDT New Technology & Application Forum (FENDT), 2021.
[7] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in: Proc. IEEE conference on computer vision and pattern recognition, 2014.
[8] R. Girshick, Fast R-CNN, in: Proc. IEEE international conference on computer vision, 2015.
[9] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, IEEE Transactions on Pattern Analysis and Machine Intelligence 39(6)(2017) 1137-1149.
[10] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in: Proc. IEEE conference on computer vision and pattern recognition, 2016.
[11] H. Law, J. Deng, Cornernet: Detecting objects as paired keypoints, in: Proc. European conference on computer vision (ECCV), 2018.
[12] X.-Y. Zhou, D.-Q. Wang, P. Krähenbühl, Objects as points. <https://arxiv.org/abs/1904.07850>, 2019 (accessed 30.04.2019).
[13] M. Li, H. Ge, H. Wang, IMG-CenterNet: An optimized algorithm based on CenterNet for pedestrian detection, in: Proc. 2022 IEEE 6th Information Technology and Mechatronics Engineering Conference (ITOEC), 2022.
[14] S. Liu, H. Ma, Combined attention mechanism and CenterNet pedestrian detection algorithm, in: Proc. 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 2021.
[15] Y. Sun, Z. Li, L. Wang, J. Zuo, L. Xu, M. Li, Automatic Detection of Vehicle Targets Based on CenterNet Model, in: Proc. 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE), 2021.

[16] X. Jin, J. Che, Y. Chen, Weed Identification Using Deep Learning and Image Processing in Vegetable Plantation, IEEE Access 9(2021) 10940-10950.