Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Method for restoring occluded boundaries of dense juvenile abalone in aquaculture environments based on discrete cosine transform

Jun Yue
Ludong University
Changyi Dai
Ludong University
Guangjie Kou
kouguangjie@126.com
Ludong University
Zihao Mao
Ludong University

Research Article

Keywords: occlusion restoration, dense small targets, deep learning, discrete cosine transform

Posted Date: September 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8027547/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Method for restoring occluded boundaries of dense juvenile abalone in aquaculture environments based on discrete cosine transform

Jun Yue,[a] Changyi Dai,[a] Guangjie Kou, [a,*] Zihao Mao[a]

[a] Ludong University, College of Computer Science and Artificial Intelligence, Yantai, China

Abstract. The detection and grading of abalone seedlings is a key step in the industrial breeding of abalone. In real aquaculture environments, abalone seedlings are small in size, densely distributed, and have complex occlusions. These factors result in low efficiency and large errors in manually calculating biomass, while existing instance segmentation based occlusion boundary restoration methods have low accuracy in processing juvenile abalone occlusion data. In response to the above issues, this paper proposes a dense juvenile abalone occlusion boundary restoration method based on discrete cosine transform. A cross scale fusing module is constructed in the network structure of this algorithm model, which enhances the expression of features by fusing features of different scales in the context, learns the adjacent scale features fused by juvenile abalone, and improves the segmentation accuracy of juvenile abalone; At the same time, a learnable noise module was designed in the algorithm model to learn the occluded boundaries of juvenile abalone by introducing noise in different batches, expanding the complexity of discrete cosine coding, learning more encoding information, improving the coherence of DCT vectors, and enhancing segmentation accuracy; Finally, the algorithm model proposed in this article was validated for its effectiveness on the AbData dataset of juvenile abalone, which includes multiple occlusion scenarios and is designed for real-world aquaculture environments.

Keywords: occlusion restoration, dense small targets, deep learning, discrete cosine transform.

* Corresponding author: Guangjie Kou, E-mail: kouguangjie@126.com

1 Introduction

Abalone is an important economic seafood with high nutritional value and broad market demand. China has a long history of abalone farming, and related technologies have gradually entered a mature stage[1-3]. Seedling selection is an important upstream industry in abalone aquaculture, which requires multiple grading and screening of juvenile abalone at this stage. The traditional manual grading and screening method has physical contact with abalone, which not only affects the

appearance of abalone, but also easily introduces unpredictable virus hazards. In addition, manual methods also have many disadvantages such as high labor costs and easy introduction of manual errors. To effectively solve the above problems, more and more people in the field of abalone farming are introducing computer vision, a non-contact detection and grading method, to improve the accuracy and efficiency of juvenile abalone grading. Faced with the complex occlusion of dense small object detection and segmentation tasks such as juvenile abalone grading, classical image processing methods[4-6] are no longer able to meet the practical needs of breeding grading. In recent years, with the widespread application of deep learning technology in the field of computer vision, a new solution has been provided for solving the above-mentioned problems.

2 Background and Existing Work

In actual aquaculture environments, abalone seedlings have varying degrees of intra class and inter class obstruction, which poses great difficulties for subsequent seedling selection and treatment. In the research of occlusion object detection methods, Ke L et al.[7] used two GCN layers to detect overlapping instances, decoupling the boundaries of obstructing and occluded objects. Jiang et al.[8] adopted a de-overlapping network to establish the interaction between the overall instance and its subregions, explicitly decoupling the two objects into intersecting regions. Ge Z et al.[9] proposed a predictive branch decoupling head to improve the convergence speed of the predictive branch decoupling head. Xiyang Dai et al.[10] proposed a novel dynamic head framework that combines object detection head with attention, achieving scale perception by synergistically combining multiple self attention mechanisms between feature layers. Rui Zhu et al.[11] proposed a pre-training model that abandons classification and trains the detection model from scratch, modifying ResNet to better adapt to the detection framework. Feng et al.[12] proposed a TOOD single-stage object detection model and separately predicted the coordinates for classification and localization. The dynamic instance segmentation algorithm based on location proposed by Wang et al.[13] effectively improves the ability of object detection in complex environments. Zhan et al.[14] used unsupervised learning to complete the task of removing occlusions by restoring the potential occlusion order and merging the invisible parts of the occluded objects. Chen et al.[15] proposed a hybrid task cascade framework that combines boundary regression and mask prediction, and utilizes spatial background to improve the performance of instance segmentation. Zhang et al.[16] designed a new aggregation loss to approach and compactly locate corresponding objects, in order to integrate

prior information into the network to handle occlusion. Ke et al.[17] proposed an information loss region detection mechanism, and the constructed quadtree structure can achieve significant performance improvement on object edges. Li et al.[18] proposed a second-order attention module to capture more comprehensive features and improve the accuracy of re identification of occluded pedestrians. Jin et al.[19] developed an occlusion aware memory refinement module to enhance feature maps for recursive lane detection in videos. Liang et al.[20] proposed a shape prior based occlusion aware segmentation method for complex scenes, which improved the accuracy of fruit contour segmentation in complex natural environments.

In terms of restoring occluded boundaries, Zhang et al.[21] combined fine-grained features and additional semantic heads as guidance to compensate for the loss of details. Wen et al.[22] improved the DCT mask by refining features through multiple independent batches, resulting in more compact encoding and improved efficiency. Liu et al.[23] proposed a new sea sand imaging method for identifying and distinguishing shells obscured by sea sand. Kibet et al.[24] proposed a YOLOv3 TensorFlow method for detecting and counting occluded abalone. Xu et al.[25] proposed a gait recognition method based on occlusion perception model, which uses a human body network model to match unobstructed body parts. Specker et al.[26] used a multi camera tracking system to perceive occluded objects and filter out false or missed targets. Bejjani et al.[27] proposed a data-driven hybrid planner for generating occlusion aware actions in a closed loop to solve the task of retrieving target objects from cluttered shelves. Zhang et al.[28] applied the SAM model for segmenting everything to medical image segmentation, but failed to achieve satisfactory results. Karimov et al.[29] achieved precise land segmentation accuracy by redesigning U-net and optimizing the loss function. Ma et al.[30] improved the deep learning algorithm of transformer, achieving more universal cell analysis and more accurate segmentation results. Sapkota et al.[31] compared the performance of YOLOV8 and mask rcnn models in complex orchard environments. Gu et al.[32] improved the convolutional neural network to efficiently and accurately measure the number of mussels. Sánchez et al.[33] obtained annotation results for instance segmentation through a weakly supervised approach that combines manual and semi-automatic annotation.

Although the above methods have achieved good processing results on their respective datasets, facing the application requirements of abalone breeding, due to the dense distribution, varied shapes, and severe occlusion between abalone seedlings, effectively detecting abalone seedlings and restoring occluded boundaries still pose great challenges. This article proposes a dense

juvenile abalone occlusion boundary restoration method based on discrete cosine transform. Firstly, a dense juvenile abalone dataset AbData containing multiple occlusion situations was constructed for real aquaculture environments, providing effective support for the training and testing of algorithm models; Secondly, a cross scale fusing module was proposed to extract image features from multiple scales, and a learnable matrix was used to reparameterize the network to improve the segmentation accuracy of juvenile abalone; Finally, a discrete cosine learnable noise module is proposed to expand the complexity of discrete cosine encoding, improve the coherence of discrete cosine vectors, and learn the occluded boundaries that are missing in images.

3 Methodology

3.1 Patch-DCT Model

Patch-DCT[23] is a method for high-quality instance segmentation that improves segmentation quality by decomposing the mask decoded from DCT vectors into multiple patches and refining each patch using classifiers and regressors. This method combines the advantages of two commonly used methods for achieving fine instance segmentation (multi-stage cascading to refine segmentation results, vector compression to preserve high-resolution features), and overcomes the negative impact caused by the relationship between each element of highly compressed feature vectors and the features of all regions in the original high-resolution feature map.

Since each pixel in the mask is calculated using the entire vector, refining the vector can cause many pixels in the mask to change, even well segmented pixels may change. Changing one element in the vector can result in significant changes to the entire mask. For this purpose, Patch-DCT divides the mask into many independent batches and refines each batch. The mask of each batch can be converted into a vector through DCT, changing the elements in a certain vector and only changing the mask in the batch corresponding to that vector. The batches of masks are divided into three categories: foreground batches (including only foreground), background batches (including only background), and mixed batches (including both foreground and background). The compression vectors of each batch have distinct characteristics, making it easier for classifiers to distinguish between these three types of patches. Of course, there may also be errors in foreground and background recognition, where there is a square hole in the middle of the foreground segmentation result or a foreground block in the background area. The model in this article is

improved by selecting Patch-DCT and learning DCT vectors to restore occluded boundaries. And combine feature maps of different sizes to optimize the phenomenon of holes.

3.2 Network Model for restoring the occluded boundary of dense juvenile abalone in aquaculture environment

This article proposes a CD-DCT dense juvenile abalone occlusion boundary restoration model based on Patch DCT, as shown in Fig.1.

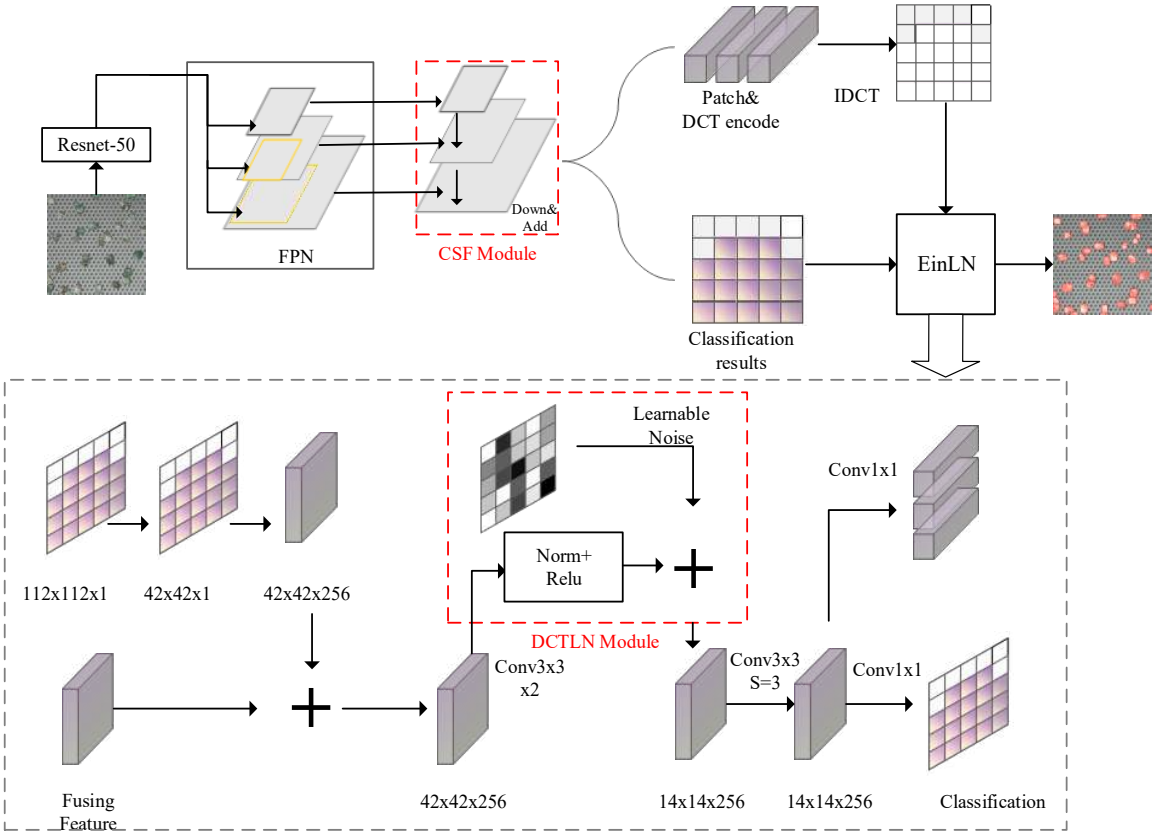


Fig. 1 Structure diagram of dense juvenile abalone occlusion boundary restoration model based on discrete cosine transform

The red box represents the module proposed in this article, which is used to improve the segmentation accuracy of abalone occlusion boundaries in severe dense and mutually occluded situations. The proposed Adjacent Region Completing Module (CSF) can learn information about abalone at different scales, which is helpful for the localization of individual abalone; The proposed Discrete Cosine Learnable Noise Module (DCTLN) can learn the discrete cosine encoding of occluded abalone images.

3.3 Cross Scale Fusing Module (CSF)

Multi scale feature fusion is a commonly used and effective method in image segmentation. But it is usually achieved through attention mechanisms or multi-layer convolution methods, which require a lot of video memory resources and have limited improvement effects. Therefore, this article designs a lightweight cross scale fusing module as shown in Fig. 2 that reduces network complexity while using as many adjacent features as possible, so that the module can significantly improve the performance of the model without increasing additional computational costs.

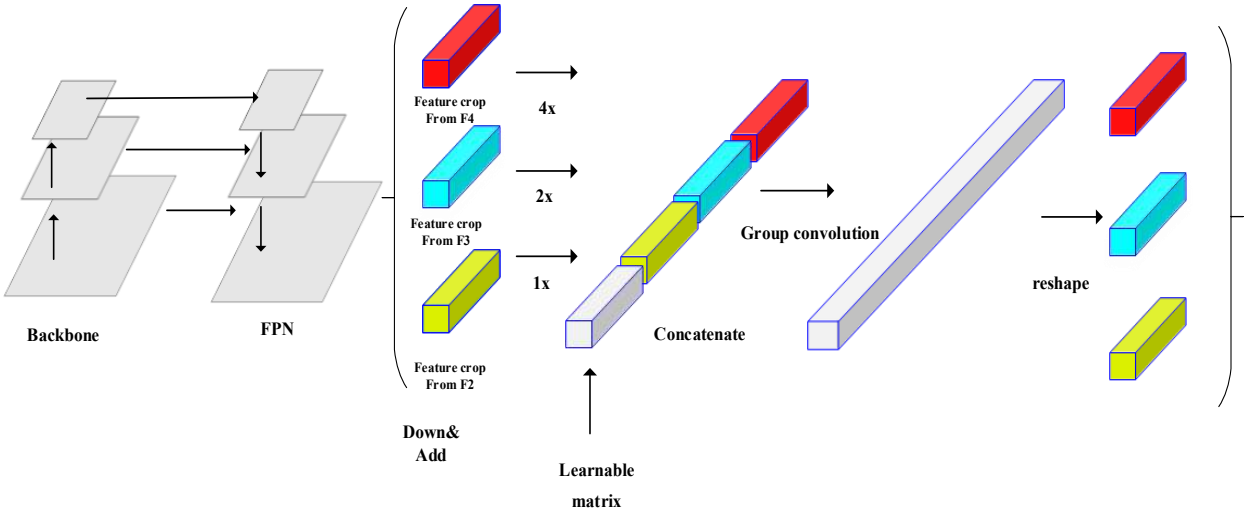


Fig. 2 Cross Scale Fusing Module (CSF)

Different from the previous methods of context fusion or element summation and splicing, the key idea of this paper is to use a learnable matrix to adaptively correct the fusion weight of feature mapping on each scale. The resolution feature of F2 is represented as x^l, and the feature of the previous layer x^n is adjusted to the same shape as x^l. Firstly, the feature channel of the current layer is compressed with a 1x1 convolution layer, and then the resolution of the previous layer is improved by interpolation to obtain an upsampled feature map of the same size, which is then fused with a layer of learnable matrix to obtain the final output feature map. The specific feature fusion method can be expressed by equation (1).

$$y_{ij}^{l} = conv(x_{ij}^{1\to l}) + x_{ij}^{2\to l} + Z_L \tag{1}$$

In equation (1), y_{ij}^{l} is the output characteristic graph of the current layer, conv represents convolution layer operation, and $x_{ij}^{1\to l}$ represents that the characteristic shape of the current layer

is aligned to the characteristic shape of layer 1 by down sampling method. ZL is a learnable zero matrix.

In convolution layer operation, k=3 is used by default, so padding=1 should be used to align the shape. Such operation can be regarded as a convolution kernel scan basis weight of 3 × 3 to extract the representation to build the required kernel. Then inspired by residual learning, this paper hopes that this part of the module can learn the increment of base weight instead of the original mapping. Therefore, ZL, a learnable zero matrix, is introduced, so that ZL can be updated in the optimizer. As mentioned in Patch-DCT, when the classifier may have classification errors, it will lead to some holes. On the one hand, because the information range is different, the use of adjacent pixel information can effectively make up for this; On the other hand, the use of learnable zero matrix can smooth features, enhance the prior of model structure, and improve the ability of feature representation.

3.4 Discrete Cosine Learnable Noise Module (DCTLN)

The frequency of image refers to the change frequency of gray value, which is the gradient of gray in plane space. The main component of the image is the low-frequency information, which forms the basic gray level of the image; The intermediate frequency information determines the basic structure of the image and forms the main edge structure of the image; High frequency information forms the edges and details of the image. For the problem of boundary segmentation in images, this paper optimizes the high frequency part. A new method of noise learning using discrete cosine is proposed, which can effectively capture the effective data in the image boundary through the global view. Firstly, the DCT-mask method is used to encode the mask to the frequency domain by dividing the mask of K × K into m × m patches.

$$\mathrm{M}^{\mathrm{f}}_{\mathrm{K}\times\mathrm{K}}(u,v)=\frac{2}{K}C(u)C(v)\sum_{z=0}^{K-1}\sum_{y=0}^{K-1}M_{\mathrm{K}\times\mathrm{K}}(x,y)\cos\frac{(2x+1)u\pi}{2K}\cos\frac{(2y+1)v\pi}{2K} \tag{2}$$

Then the n-dimensional DCT vector is obtained by selecting the highest n elements through discrete cosine vector regression. Each batch is coded independently to avoid the interference between vectors, which is used to extract more distinguishing features. Finally, features and masks are fused through a noise matrix that can update parameters, and batchnormal is used to facilitate training and relu function is used for connection. In order to reduce the complexity, Einstein matrix

multiplication (EMM) is used to obtain an efficient block diagonal matrix. The specific equation is shown in equation (3).

$$\mathrm{M} = \mathrm{EMM}(\mathrm{Relu}(\mathrm{Norm}(\mathrm{M}^{'}))\mathrm{Z}_{\mathrm{N}}) \tag{3}$$

In equation (3), M' is the fused feature, Relu represents relu function operation, Norm represents batchnormal function operation, and ZN is the randomly generated learnable noise matrix. The loss function of mask is defined as equation (4). Where λ_0 and λ_s are loss weights, and the first term L_{dctN} is defined to predict the loss of the n-dimensional vector of the entire mask. In equation (5), $\hat{V}_i$ and V_i are the i th element of the truth vector and prediction vector respectively, R is the loss function, n is the length of the vector, and classification loss $L^s_{clspadct}$ is the cross entropy. Regression loss L^s_{dctn} is shown in equation (6). Where Nm is the number of mixed patches and n is the length of the vector. If the kth patch is a mixed patch, pk is 1, otherwise pk, and only the DCT vector of the mixed patch will be regressed.

$$\mathrm{L}_{mask} = \lambda_0 L_{dctN} + \lambda_s (L^s_{clspadct} + L^s_{dctn}) \tag{4}$$

$$L_{dctN} = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}}^{N} R(\hat{V}_i - V_i) \tag{5}$$

$$L^s_{dctn} = \frac{1}{\mathrm{N_m}} [p^k (\frac{1}{\mathrm{n}} \sum_{\mathrm{i}}^{n} R(\hat{V}_i - V_i))] \tag{6}$$

The design motivation of DCTLN is that the images in the AbData dataset have synthetic boundaries. There is a natural difference between the natural boundary and the synthetic boundary, that is, the continuity of the boundary. The discontinuity in the synthetic data leads to the fluctuation on the subtle boundary of the prediction. Therefore, by introducing a learnable noise matrix to adjust different boundary information, the expression ability of DCT can be enhanced. On the other hand, the boundary of the occluded object is missing from the image. Only the boundary of the occluded object is complete, and the design of DCTLN makes up for the missing boundary DCT vector. It makes the boundary of the occluded object tend to be complete after adding the DCT vector.

4 Analysis of experimental results

4.1 experimental parameter setting

The experiment in this paper is based on Pytoch framework and Detectron2[34] platform to realize the network, and the software and hardware platform configuration parameters used are shown in Table 1.

Table 1 Software and Hardware Configuration Parameters

Equipment	Configuration
Operating System	Ubuntu20.04
CPU	Intel® Core™ i7-12700KF
GPU	NVIDIA GeForce RTX 3080TI
GPU Computing Platform	Cuda11.3,Cudnn8.8
Deep Learning Framework	Pytorch
Compiler	Anaconda，Vscode
Programming Language	Python3.7

The self built data set AbData is adopted. In order to meet the application requirements in the aquaculture environment, the data set includes 148540 data sets of dense shielded abalone seedlings under the complex environment of seaweed and seaweed shelter. It is divided into training set (118832), test set (14854) and verification set (14854) according to the ratio of 8:1:1. Each image size is 320x320 and the maximum iteration 4000 is carried out under the condition of batchsize 2. The warmup is used as the learning rate, which makes the model tend to be stable under the small learning rate. Then the initial learning rate of 0.002 is selected for training, which makes the convergence speed of the model faster and the model effect better. The AbData dataset is in CoCo format. The network is trained in the training set, and the results are evaluated on the validation set and the test set. The AP index, APS index and FPS index commonly used in the segmentation task are used to evaluate the performance of the model.

4.2 Ablation Experiment and result analysis

In order to verify the effectiveness of the two modules proposed in this paper, the benchmark model, the network model with discrete cosine learnable noise module (DCTLN), the network model with cross scale fusing module(CSF) and the network model with both modules (CD-DCT)

were tested in the AbData dataset using consistent hyper parameters. The test results are shown in Table 2.

Table 2 Ablation Experiment

Base	CSF	DCTLN	AP	AP50	AP75	APs	APm	FPS
1			57.882	87.565	65.218	57.706	67.693	23.6
2	√		57.911	87.471	65.147	57.800	66.744	25.3
3		√	58.344	87.550	65.990	58.206	67.664	21.9
4	√	√	58.789	87.600	66.239	58.545	68.218	25.2

The experimental results show that APS is improved by 0.1% after adding the adjacent region completion module. The adjacent area completion module is a module that focuses on boundaries, which focuses on generating more accurate boundaries. By introducing additional learnable matrix to adjust the weight, each convolution kernel can learn more different features. This method can flexibly respond to the input features without consuming too much computing resources. After the module expands the boundary details, it can obtain a certain improvement in accuracy and speed.

After adding the (DCTLN module, because the addition of a learnable matrix can enrich the DCT vector, the calculation amount is slightly increased and the speed is reduced acceptably. It can be seen from table 2 that the addition of this module has improved AP by 0.35%, which indicates that rich vectors help to further refine the boundary.

4.3 Comparative Experiment and Result Analysis

In order to verify the performance of the proposed CD-DCT, the CD-DCT model is compared with the experimental results of YOLOx[9], Solov2[13], Transfiner[17], Patch-DCT[22]in the AbData dataset and the same operating environment, and the experimental comparison results are shown in Table 3. The proposed CD-DCT algorithm has better performance in accuracy, which is improved by 0.91% AP compared with the benchmark model, and 1.50% AP compared with the Transfiner. The difference of AP shows the effectiveness of this algorithm in generating higher quality masks.

The data in this paper are mainly small targets, and APs index can better reflect the effectiveness of the algorithm. Compared with Transfiner, APs increased by 1.31% and 0.84% compared with

the benchmark Patch-DCT, which corresponds to the increase of AP. Secondly, in terms of running speed, the method in this paper is 1.6 higher than that of the benchmark model in FPS, which is an improvement obtained by CSF module in learning boundary parameters. The running speed of the method in this paper is more than 4 times higher than that of the Transfiner, which is due to the need for refined learning of 112×112 outputs in the Transfiner, while in Patch-DCT, due to batch learning, only 14×14×6 outputs need to be learned, which greatly reduces the complexity of training and reasoning.

Table 3 Comparison Experiments of Different Algorithms

Method	**AP**	**AP50**	**AP75**	**APs**	**APm**	**FPS**
YOLOx(2021)	37.8	82.8	30.4	39.6	28.0	6.2
Solov2（2020）	36.4	73.5	32.6	36.4	42.0	4.7
Transfiner（2022）	57.290	**87.664**	65.377	57.233	64.739	4.6
Patch-DCT（2023）	57.882	87.565	65.218	57.706	67.693	23.6
CD-DCT（our）	**58.789**	87.600	**66.239**	**58.545**	**68.218**	**25.2**

In Fig.3, this paper visualizes some results of the CD-DCT algorithm on the AbData dataset. In the results of the first column, there are two different instances in the lower left corner, which can be distinguished by the model in this paper. Although other models can correctly restore their boundaries, Patch-DCT identifies them as one, Transfiner identifies three instances, which are false detected, and Solov2 identifies one, which is missing detected. In the case where the left side of the second column is covered by water plants, the boundary that can be restored in this model is more appropriate, and the matching area score of other models is lower. In the results of the third column, the model in this paper can effectively restore the boundary of the two overlapping instances in the lower left corner, while Solov2 and Patch-DCT are detected as one, and there is a phenomenon of missing detection. Therefore, the CD-DCT algorithm proposed in this paper can achieve better occlusion boundary recovery and has advantages in speed under the premise of small size, dense individual distribution and serious occlusion within the category of juvenile abalone, and other heterogeneous occlusion in aquatic and grass breeding environment.

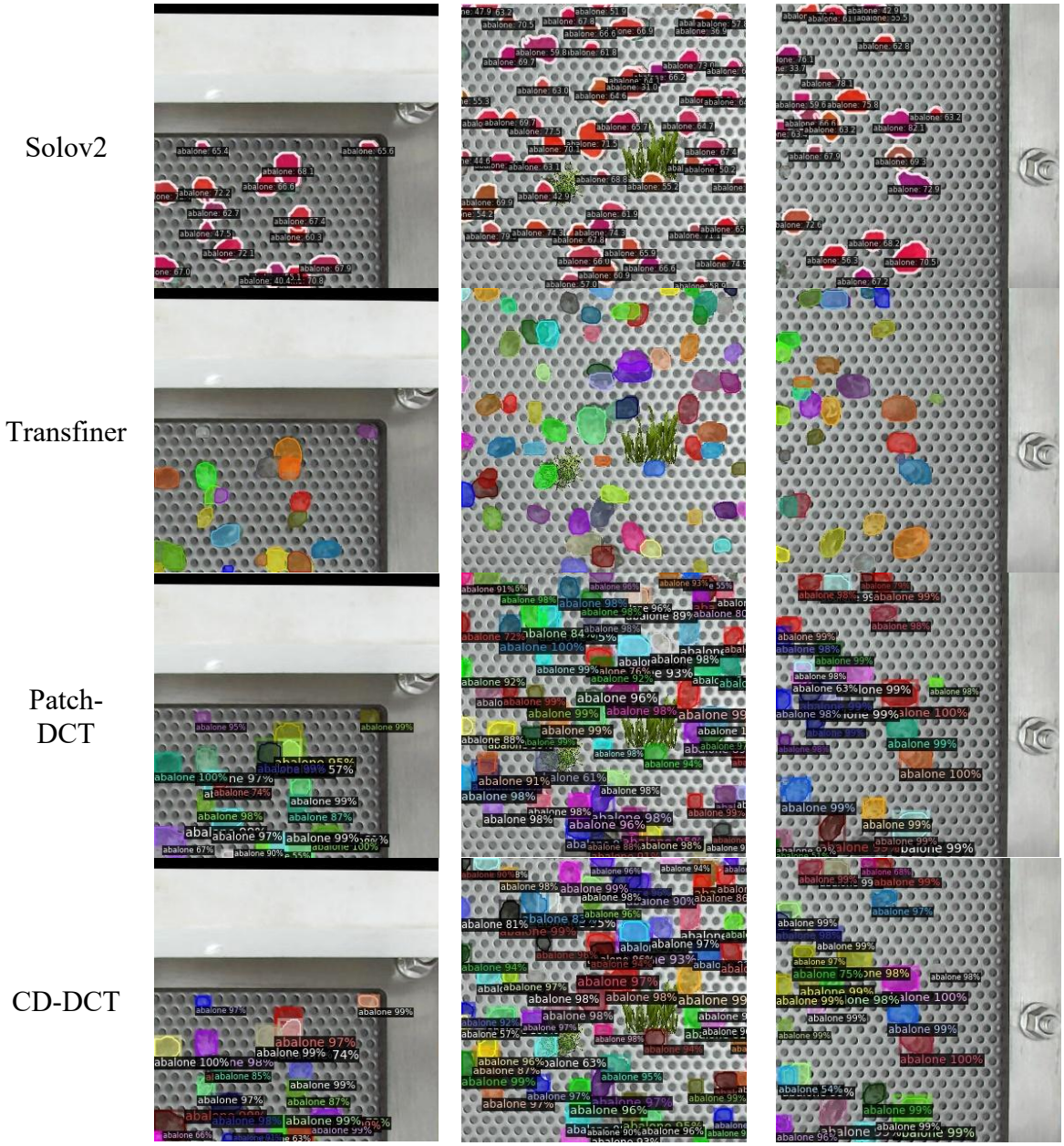


Fig. 3 Comparison of visualization results

5 Conculsion

Based on Patch-DCT model and real abalone culture environment, this paper proposes an occlusion boundary recovery method for dense occlusion of juvenile abalone. The innovations of this method are as follows: firstly, an cross scale fusing module is designed, which integrates the effective information of the region, enhances the prior of the model structure, and solves the cavity phenomenon mentioned in the original text; Secondly, the discrete cosine learnable noise module is designed to enrich the DCT vector to make the boundary smooth and coherent, enhance the

expression of boundary features, so as to improve the segmentation accuracy of the occluded boundary of juvenile abalone, and solve the problem of poor segmentation quality caused by the difference between the natural boundary and the synthetic boundary. Finally, the effectiveness of the proposed algorithm is verified on the AbData dataset. The proposed algorithm model can effectively improve the accuracy of occlusion recovery of juvenile abalone, and lay a technical foundation for the subsequent study of juvenile abalone vision.

Author contributions Jun Yue and Guangjie Kou designed and executed the experiments, analyzed the data, and written the manuscript. Changyi Dai prepared the data set, executed the experiments and written the manuscript. Zihao Mao contributed to thorough review and editing of the manuscript. All authors gave approval to the final version of the paper.

Funding Shandong Province Key R&D Program Project (2022TZXD005), Shandong Province Science and Technology SMEs Innovation Capacity Enhancement Project (2021TSGC1003), Yantai City Key R&D Program Project (2022XCZX079).

Data availability Data will be made available from corresponding author on reasonable request.

Decalarations

Competing interest The authors declare no competing interests.

References

1. Gao X, Zhang M, Luo X, et al., “Transitions, challenges and trends in China's abalone culture industry,” Reviews in Aquaculture 15(4), 1274-1293 (2023)
2. Wu F, Que H, et al., “History, current situation and future development trend of abalone culture by bottom sowing in China,” Marine Sciences 44(8): 56-68 (2020)
3. Luan M, Yang L,et al., “Study on the culture technology of abalone haliotis discus hannai in the deep sea,” Scientific Fish Farming (05): 66-67 (2023)
4. Gao W, Zhang X, Yang L, et al., “An improved Sobel edge detection,” 2010 3rd International conference on computer science and information technology. IEEE pp.67-71 (2010)
5. Zhang Y J., “An overview of image and video segmentation in the last 40 years,” Advances in Image and Video Segmentation 2006: 1-16 (2006)

6. Zou L , Chen Q , Wu Z ,et al., “A level-set method for fast image segmentation based on local pre-fitting and bilateral filtering,” Engineering Computations 42 (1): 96–116 (2025)
7. Ke L, Tai Y W, Tang C K., “Deep occlusion-aware instance segmentation with overlap bilayers,” Proceedings of the IEEE/CVF conference on computer vision and pattern recognition pp.4019-4028 (2021)
8. Jiang H, Zhang R, Zhou Y, et al., “Donet: Deep de-overlap** network for cytology instance segmentation,” Proceedings of the IEEE/CVF conference on computer vision and pattern recognition pp.15641-15650 (2023)
9. Ge Z., “Yolox: Exceeding yolo series in 2021,” arxiv preprint arxiv:2107.08430 (2021)
10. Dai X, Chen Y, et al., “Dynamic head: Unifying object detection heads with attentions,” Proceedings of the IEEE/CVF conference on computer vision and pattern recognition pp.7373-7382 (2021)
11. Zhu R, Zhang S, Wang X, et al., “ScratchDet: Training single-shot object detectors from scratch,” Proceedings of the IEEE/CVF conference on computer vision and pattern recognition pp.2268-2277 (2019)
12. Feng C, Zhong Y, Gao Y, et al., “Tood: Task-aligned one-stage object detection,” 2021 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE Computer Society pp.3490-3499 (2021)
13. Wang, X.; Zhang, R.; Kong, T.; Li, L.; Shen, C., “Solov2: Dynamic and fast instance segmentation,” Adv. Neural Inf. Process. Syst. 33:17721–17732 (2020)
14. Zhan X, Pan X, Dai B, et al., “Self-supervised scene de-occlusion,” Proceedings of the IEEE/CVF conference on computer vision and pattern recognition pp.3784-3792 (2020)
15. Chen K, Pang J, Wang J, et al., “Hybrid task cascade for instance segmentation,” Proceedings of the IEEE/CVF conference on computer vision and pattern recognition pp.4974-4983 (2019)
16. Zhang S, Wen L, Bian X, et al., “Occlusion-aware R-CNN: Detecting pedestrians in a crowd,” Proceedings of the European conference on computer vision (ECCV) pp.637-653 (2018)
17. Ke L, Danelljan M, Li X, et al., “Mask Transfiner for high-quality instance segmentation,” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition pp.4412-4421 (2022)
18. Li Y, Liu Y, Zhang H, et al., “Occlusion-Aware Transformer With Second-Order Attention for Person Re-Identification,” IEEE Transactions on Image Processing 33(000):3200-3211 (2024)
19. Jin D, Kim C S., “OMR: Occlusion-Aware Memory-Based Refinement for Video Lane Detection,” European Conference on Computer Vision. Springer, Cham pp.129-145 (2025)
20. Liang J, Huang K, Lei H, et al., “Occlusion-aware fruit segmentation in complex natural environments under shape prior,” Computers and Electronics in Agriculture 217:108620 (2024)

21. Zhang G, Lu X, Tan J, et al., “Refinemask: Towards high-quality instance segmentation with fine-grained features,” Proceedings of the IEEE/CVF conference on computer vision and pattern recognition pp.6861-6869 (2021)
22. Wen Q, Yang J, Yang X, et al., “Patchdct: Patch refinement for high quality instance segmentation,” The Eleventh International Conference on Learning Representations pp.1-15 (2023)
23. Liu T, Ju Y, Lyu H, et al., “Identification and analysis of seashells in sea sand using computer vision and machine learning,” Case Studies in Construction Materials 18, e02121 (2023)
24. Kibet D, Shin J H., “Counting Abalone with High Precision Using YOLOv3 and DeepSORT,” Processes 11(8): 2351 (2023)
25. Xu C, Makihara Y, Li X, et al., “Occlusion-aware human mesh model-based gait recognition,” IEEE transactions on information forensics and security 18: 1309-1321 (2023)
26. Specker A, Stadler D, Florin L, et al., “An occlusion-aware multi-target multi-camera tracking system,” Proceedings of the IEEE/CVF conference on computer vision and pattern recognition pp.4173-4182 (2021)
27. Bejjani W, Agboh W C, Dogar M R, et al., “Occlusion-aware search for object retrieval in clutter,” 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) 4678-4685 (2021)
28. Zhang Y, Shen Z, Jiao R., “Segment anything model for medical image segmentation: Current applications and future directions,” Computers in Biology and Medicine 2024:108238 (2024)
29. Karimov S, Sotvoldiyeva D, Khalilov D, et al., “Deep neural network for semantic segmentation of satellite images,” E3S Web of Conferences. EDP Sciences pp.03006 (2024)
30. Ma J, Xie R, et al., “The multimodality cell segmentation challenge: toward universal solutions,” Nature methods 21(6):20 (2024)
31. Sapkota R, Ahmed D, Karkee M., “Comparing YOLOv8 and Mask R-CNN for instance segmentation in complex orchard environments,” Artificial Intelligence in Agriculture 13: 84-99 (2024)
32. Gu Y F, Hu J, Williams G A., “A comparative study on CNN-based semantic segmentation of intertidal mussel beds,” Ecological Informatics 75: 102116 (2023)
33. Sánchez J S, Lisani J L, Catalán I A, et al., “Leveraging bounding box annotations for fish Segmentation in underwater images,” IEEE Access 11: 125984-125994 (2023)
34. Tan M, Pang R, Le Q V., “Efficientdet:Scalable and efficient object detection,” Proceedings of the IEEE/CVF conference on computer vision and pattern recognition pp.10781-10790 (2020)

Jun Yue received the Ph.D degree in management science and engineering from China Agricultural University, Beijing, China. She is now a professor in the College of Computer Science and Artificial Intelligence, Ludong University, Yantai, China. Her research interests include computer vision and deep learning.

Guangjie Kou received the Ph.D. degree in computer science and technology from the Institute of Computing Technology, Chinese Academy of Sciences (CAS), Beijing, China. He is now a professor in the College of Computer Science and Artificial Intelligence, Ludong University, Yantai, China. His research interests include computer vision and bioinspired intelligent computing.