The highlight has been updated to the following points:

1. The abnormal behavior fishes in RAS were detected and tracked in real time.

2. Problems of stacking, deformation, occlusion and small object detection are solved.

3. The AP50 of the improved YOLOX reached 98. 4%, and the AP50:95 increased by 16. 2%.

4. Bytetrack is used to track the abnormal behavior for avoiding the re-identification.

1. In the complex scenes of RAS, MOTA and IDF1 can reach more than 95%.

List of changes in the main text: (All of these changes are highlighted in the main text. )

1. In Section 2. 1 on page 7, the context “The abnormal behavior dataset collected are mainly the turning-over behavior of *Larimichthys crocea* caused by environmental discomfort, bacterial diseases and parasites. The dataset acquisition devices used in the experiments are a laptop and an Intel D455 depth camera. ” has been added.
2. In page 9，the context“The main function of the attention mechanism is to make the neural network pay attention to the needed information, .... Some of them are effective in feature extraction. Some combine different features to improve the accuracy of feature extraction.”has been added .
3. In page 11, the context “Intersection over [Union](https://so.csdn.net/so/search?q=Union&spm=1001.2101.3001.7020" \t "https://blog.csdn.net/weixin_48167570/article/details/_blank):Overlapping rate of bounding box and ground truth” has been added to explain the abbreviation of IOU.
4. In page 11, the context “The division of the data set is shown in Table 1, in which there are 5986 images in the training set, 728 images in the test set, and 521 images in the validation set.” has been added.
5. In page 13, the context “ Fig. 5 is a PR curve. The closer the curve is to the upper right, the better the performance of the detector...., which proves that the proposed improvement is effective and reasonable.” has been added to describe the PR curve..
6. In page 13，the context“Regarding the detection speed of the detector, due to the improvement of YOLOX-S in this paper, although the frame rate is only 50. 67, but it can fully meet the requirements of real-time detection with higher accuracy. ”has been added to describe the detector frame rate.
7. In page 13-14 , the context “Fig. 6 shows the comparison of detection performance after adding different attention mechanisms. ..., the objects missed by other detectors can be completely detected by using YOLOX-S with CA.” has been added to explain the detection performance of the different attention mechanisms added in the Fig. 6.
8. In page 16，the context“ When tracking abnormal behavior, the main parameters of Bytetrack are set as follows: detection threshold is 0.4, ..., NN\_BUDGET (save the number of successfully matched features) is 100.” has been added.
9. In page 17-18，The tracking experiment results has been renew as:“ In the **Videos S1-S6** of the Supplementary information, different trackers are shown respectively to track the abnormal behavior of Larimichthys crocea in different complexity scenes in the actual RAS. ..., Bytetrack that avoids re-identification of appearance features is more suitable for the tracking of Larimichthys crocea. ”
10. In page 19-20，The conclusions of Bytetrack has been renew as:“For object tracking, Bytetrack is used to track the detected objects to avoid the problem of similar appearance features of fishes, . . . It can also maintain high frame rate operation, which can fully meet the effect of real-time monitoring. ”
11. In page 23，two additional citations were added as follows:

“ [Hu, J., Shen, L., Albanie, S., Sun, G., Wu, E. h. 2017,Squeeze-and-Excitation Networks. arXiv:1709.01507v4.](https://arxiv.org/abs/1709.01507v4)

Woo, [S., Park, J., Lee, J-Y , Kweon,I. S. 2018,CBAM: Convolutional Block Attention Module. arXiv:1807.06521.](https://arxiv.org/abs/1807.06521) ”

1. In page 24, the Neck section in “ **Fig. 1** ” has been modified.
2. In page 29, In “ **Fig. 4** ”, YOLOX-S has been used to replace YOLO-X and improved YOLO-X.
3. In page 30, the “ **Fig. 5** ” has been added. Correspondingly, it is used to describe recall and precision in Table 1.
4. In page 31, the “ **Fig. 6** ” has been added.
5. In page 36, the “ **Table 1** ” has been added to explain the division of datasets.
6. In page 37, “ FPS ” indicators are added in “**Table 2**”.
7. In page 38, the tracking performance results of two video sequences have been added in “**Table 3**”.

List of changes in the **Supplementary Information**: (All of these changes are highlighted in the Supplementary Information. )

1. Added 6 videos in the supplementary information. They are respectively the performance demonstration of tracking three different video sequences using the improved YOLOX-S+ Bytetrack and the improved YOLOX-S + Deepsort.

**Response to the Reviewer #1:**

**Reviewer #1:** *The author uses the improved YOLOX-S and Bytetrack to detect and track the abnormal behavior of Larimichthys crocea in the recirculating aquaculture system. The AP50:95 of the improved detector increased by 16. 2%; The MOTA and IDF1 of tracker reached 95. 02% and 97. 7%, respectively. This work is worthy of recognition and can help the automation of recirculating aquaculture. After careful consideration, I think this manuscript is suitable to be published on Aquaculture, but some modifications should be made. The details are as follows:*

We appreciate the Reviewer#1 for reviewing our paper, and also for the careful reviews and comments, which are very important for us to further improve our work and the quality of the manuscript. All the Reviewer’s questions have been rigorously concerned and revised to make a clear description and detailed explanation of our work. The itemized responses to the Reviewer’s comments are enclosed below.

***Point-by-point response to the comments:***

*Reviewer wrote:*

1. *Only the parameters of the detector were mentioned, but not the tracker. Please add the relevant parameters of the tracker.*

*Our response:*

Your comments are really thoughtful. The Bytetrack used in this paper is a tracking algorithm that is not based on deep learning, and compared with the classic deep learning-based algorithm Deepsort, but the relevant parameters of the tracker are not highlighted, which makes the data less reliable. In view of your comments, we have added relevant parameters in the page 14-15. “ When tracking abnormal behavior, the main parameters of Bytetrack are set as follows: detection threshold is 0.4, track threshold is 0.6, track buffer is 300, and matching thresh is 0.8. Since Deepsort is a deep learning-based multi-target tracking algorithm, additional training is required for this. The abnormal behavior dataset of Larimichthys crocea in RAS is trained for 40 epochs with batch size of 64, an initial learning rate of 0.01, a momentum of 0.9, and a decay of 0.0005. The main parameters of Deepsort are set as follows: detection threshold is 0.4, N\_INIT (the number of frames to keep track) is 6, NN\_BUDGET (save the number of successfully matched features) is 100. ” has been added to describe the relevant parameters of tracking.

1. *In the improved structure presentation of Fig. 4, YOLOX-S used in the article should be used instead of YOLO-X in the figure.*

*Our response:*

We are sorry for the mistake caused by our carelessness, and thank the reviewers again for carefully reviewing this paper. The error in Fig. 4 mentioned in the comments has been corrected on page 29 of the paper.

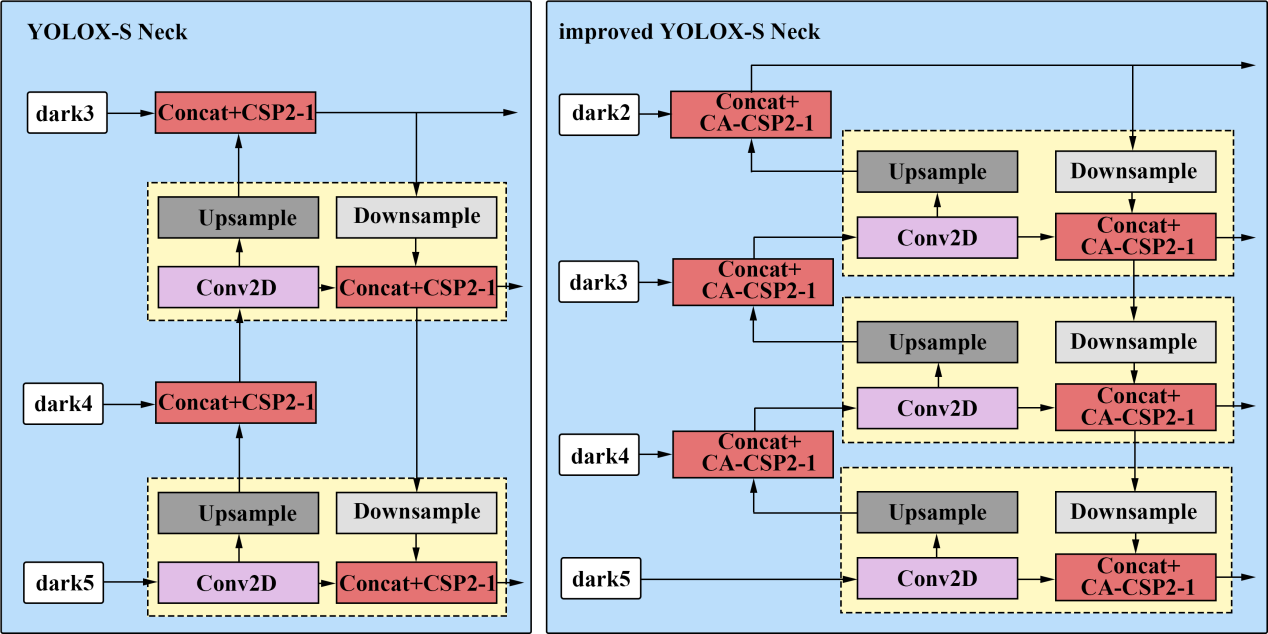


Fig. 4. Neck part structure comparison. On the left is the original Neck structure. On the right is the Neck structure using the CA-CSP module and modifying the connection.

*3. Fig. 6 shows the performance of different trackers using the same video sequence, which clearly shows the disadvantages of using appearance feature re-identification. But how many video sequences are used for the data in Table 2? If only one video sequence is used, the data is not representative, and the tracking results of different video sequences should be added, and the performance of different video sequences should be increased.*

*Our response:*

Thanks to the Reviewer#1 for their careful and thorough review, and pointed out a key problem of the article. In this paper, only a video sequence of a complex scene in a recirculating aquaculture system is used for tracking tests. Thanks for the reviewer's critical comments, which made us realize that the relevant data of only one video is not enough to support the reliability of the article. Therefore, other video sequence tracking indicators and supplementary information are added in Table 3 to improve the credibility of the article.

**Table 3** Comparison of tracking algorithm results.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Video sequence | Video | Tracker | MOTA(↑) | IDF1(↑) | ID  Switch(↓) | FPS(↑) |
| Sequence 1 | Video S1 | DeepSort | 73.03% | 61.80% | 56 | 12. 1 |
| Video S2 | Bytetrack | 95.02% | 97.70% | 1 | 39. 4 |
| Sequence 2 | Video S3 | DeepSort | 95.90% | 69.99% | 12 | 14. 8 |
| Video S4 | Bytetrack | 99.40% | 96.41% | 0 | 40. 2 |
| Sequence 3 | Video S5 | DeepSort | 97.73% | 73.24% | 22 | 12. 7 |
| Video S6 | Bytetrack | 99.10% | 99.53% | 0 | 39. 6 |

**Response to the Reviewer 2:**

**Reviewer #2：***The topic of this paper is very interesting and also practiable in RAS. I have some comments:*

We would like to thank reviewer #2 for his affirmation of the theme of this paper, as well as for the careful review of this paper and his comprehensive comments, which will be of great help and encouragement to our further research in the future. We have responded to all the reviewers' comments one by one, and the responses are as follows.

***Point-by-point response to the comments:***

*Reviewer wrote:*

*1. Materials: How are the materials used in the research obtained, how is the experimental acquisition device designed, what are the manifestations of abnormal behavior, and how are the training, testing and validation for abnormal fish divided?*

*Our response:*

We apologize for not explicitly describing the collection of materials related to fish with abnormal behavior in the article. So here we respond to your comments and add content to the article.

1. The materials used in this paper are collected from the *Larimichthys crocea* in the recirculating aquaculture system of Zhejiang Zhoushan Fisheries Research Institute.
2. The material acquisition equipment used in the experiment mainly includes a laptop and an Intel D455 depth camera. The laptop is used to connect the camera to capture video of the Larimichthys crocea in the RAS breeding tank.The abnormal behaviors detected and tracked in this paper are mainly the overturning behavior of *Larimichthys crocea* caused by environmental discomfort, bacterial diseases and parasites. “ The abnormal behavior dataset collected are mainly the turning-over behavior of *Larimichthys crocea* caused by environmental discomfort, bacterial diseases and parasites. The dataset acquisition devices used in the experiments are a laptop and an Intel D455 depth camera.” has been added in section 2.1 on page 7.
3. According to the suggestions of the reviewer, “ **Table 1** ” is added on page 36 to indicate how the datasets is divided.

**Table 1** Number of abnormal behavior fish datasets.

|  |  |  |  |
| --- | --- | --- | --- |
| Training sets | Validation sets | Test sets | Total |
| 5896 | 655 | 728 | 7279 |

*Detection: What are the advantages of Coordination Attention in this study compared to other methods? And is the base model selection with greater weights more accurate and more satisfying?*

*Our response:*

In order to reflect the advantages of the improved method proposed in this paper, the content " The main function of the attention mechanism is to make the neural network pay attention to the needed information, reduce the attention of other information and even filter the unimportant information. At present, there are many types of Attention mechanisms, such as Squeeze-and-Excitation (SE) Network ([Hu et al., 2017](#A27)), Convolutional Block Attention Module(CBAM) ([Woo et al., 2018](#A28)) etc. Some of them are effective in feature extraction. Some combine different features to improve the accuracy of feature extraction. " is added in Section 2.2.2 of this paper.

In addition, **Fig. 6** is added to show the comparison of detection heat map with different attention mechanisms, and the context " Fig. 6 shows the comparison of detection performance after adding different attention mechanisms. ... Therefore, the objects missed by other detectors can be completely detected by using YOLOX-S with CA. " is added in Section 3.1 of the paper.

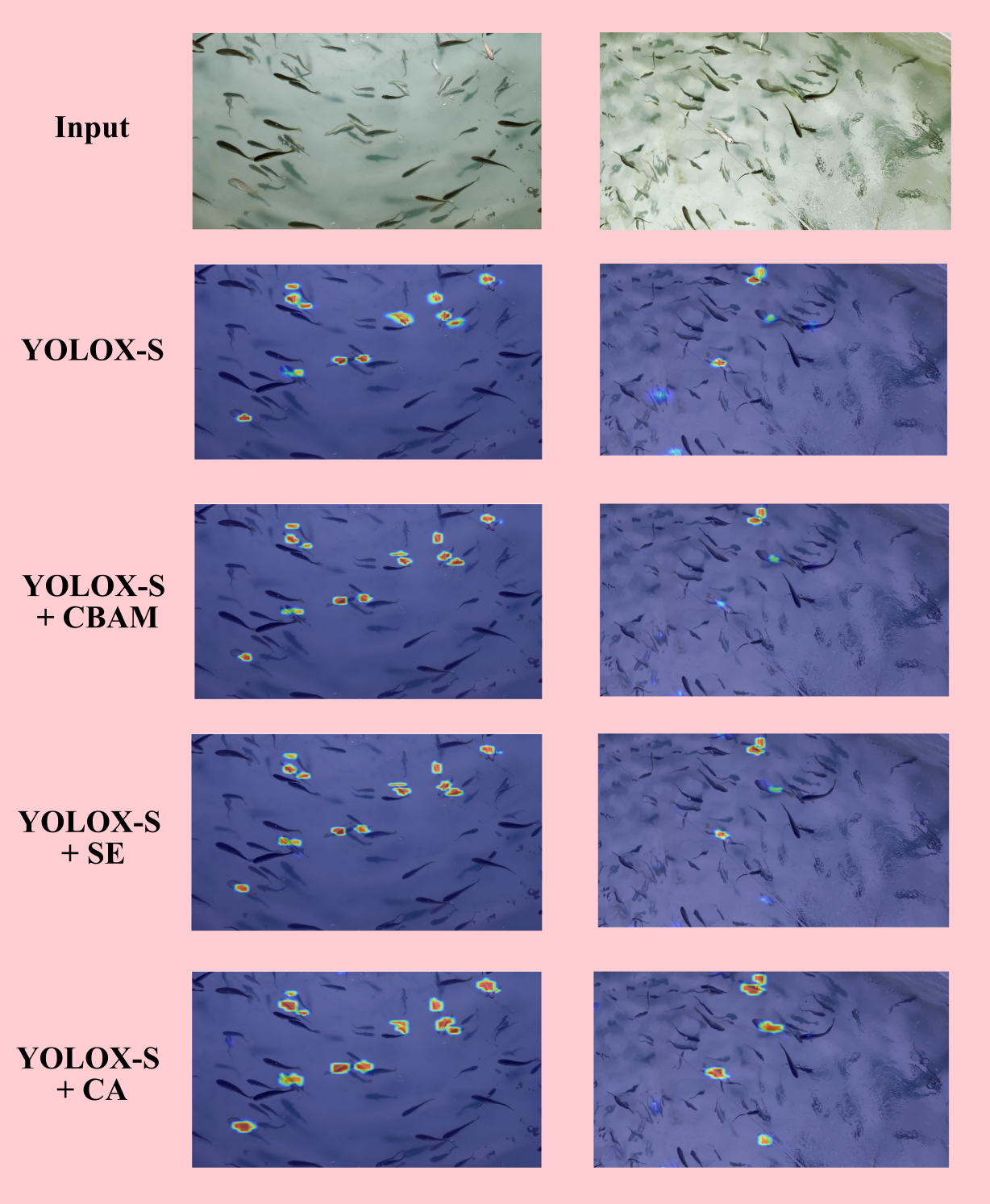


Fig. 6. Heat map comparison of detection performance of YOLOX-S with different attention mechanisms added.

Regarding the choice of model weights , there is no doubt that in terms of detection, deeper and wider networks can improve detection accuracy. However, considering that the application scenario of this topic is the aquaculture workshop of the actual recirculating aquaculture system, and the detection object is fish, a larger weight may improve the detection accuracy, But just used to detect the abnormal behavior of *Larimichthys crocea*, can lead to calculate power waste and greatly reduce the detection efficiency. Therefore, the YOLOX-S with the smallest structure depth in the standard network is selected for improvement. At the same time, the number of parameters in the network structure will be increased without affecting the detection efficiency.

*Tracking: A tracking method is described in the introduction, is it possible to compare the algorithm with the tracking algorithm in this paper? Some abbreviations, such as IOU could give a specific explanation.*

*Our response:*

Thank you for your careful review of this article. In the tracking field, there are two fields: single object tracking and multi-object tracking. The tracking algorithm SiamRPN++ is mentioned in the introduction of this paper. This algorithm is a single object tracking algorithm based on deep learning, which seems to only detect the difference in the number of objects, but their common methods are actually completely different. In terms of duration, single-object tracking focuses more on short time image sequences, while multi-object tracking generally deals with longer video, which involves the appearance, occlusion and departure of each object. Due to different tracking concepts, the mentioned algorithms are not used for comparison in this paper. In addition, the introduction mentioned that the SiamRPN++ used to achieve multi object tracking can only repeatedly send video sequences to SiamRPN++ for tracking until all object tracking is completed. Wang et.al [1] proposed a method to improve the underwater tracking of SiamRPN++ . The example in the article also shows that the algorithm can only track one object in a single run.Therefore, in the actual RAS, SiamRPN++ is not suitable for tracking *Larimichthys croceas* with abnormal behavior, and cannot be effectively used in RAS. No matter how many iterations are executed, the single object tracking algorithm cannot guarantee the ID switching of the object to be tracked and the effectiveness of practical application.

1. [Wang, Z. ,Wang, J ., Fan,R. 2021, An Underwater Single Target Tracking Method Using SiamRPN++ Based on Inverted Residual Bottleneck Block, in IEEE Access, vol. 9, pp. 25148-25157.](doi: 10.1109/ACCESS.2021.3056105)

*-Training:* *This study is only for image detection, is there any training for detection on video, as this work is hopeful to meet the actual aquaculture in the future?*

*Our response:*

Thanks to reviewer #2 for the affirmation that this research can play a role in future actual aquaculture. In practical training, video can not be used for direct training. First, the acquired video material is divided into pictures, which are annotated, and finally sent to the neural network for training. Although image training is used in this study, video sequences can be directly input to complete detection.

*-Figure: Is the concat part of the Neck layer network correctly drawn in Figure 1? Model comparison can be visualized with PR diagram.*

*Our response:*

Thanks to reviewer#2 for pointing out the mistakes in the picture and giving good suggestions. The content of this commentary has been revised and added to the article.

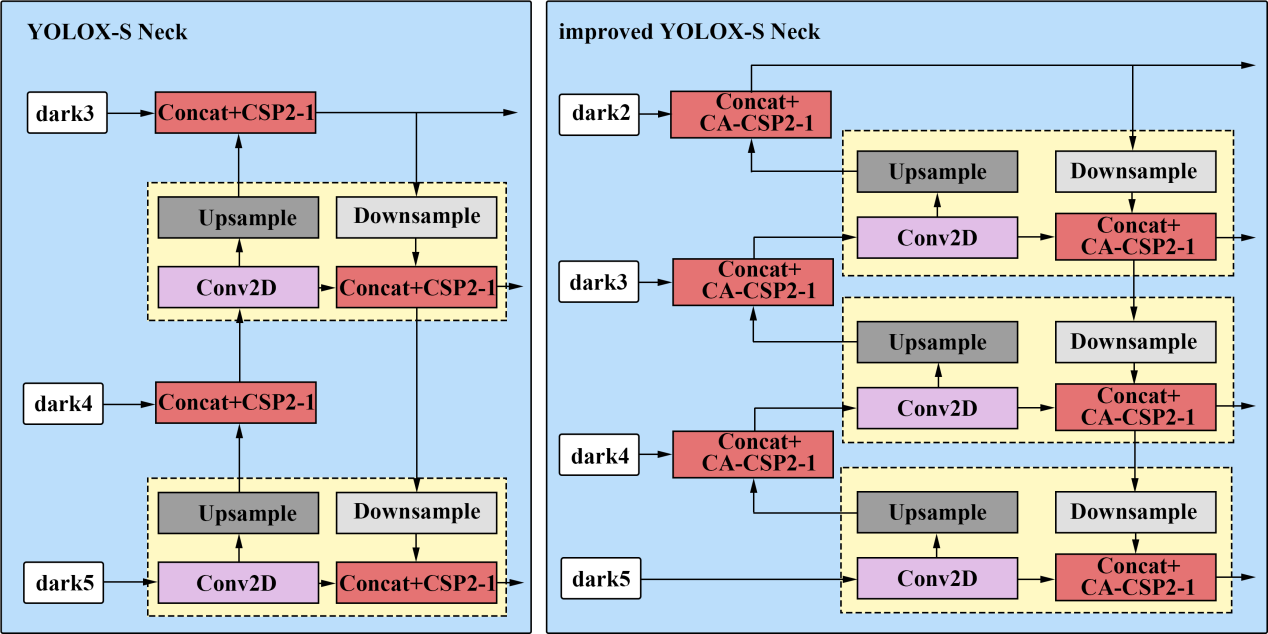


Fig. 4. Neck part structure comparison. On the left is the original Neck structure. On the right is the Neck structure using the CA-CSP module and modifying the connection.

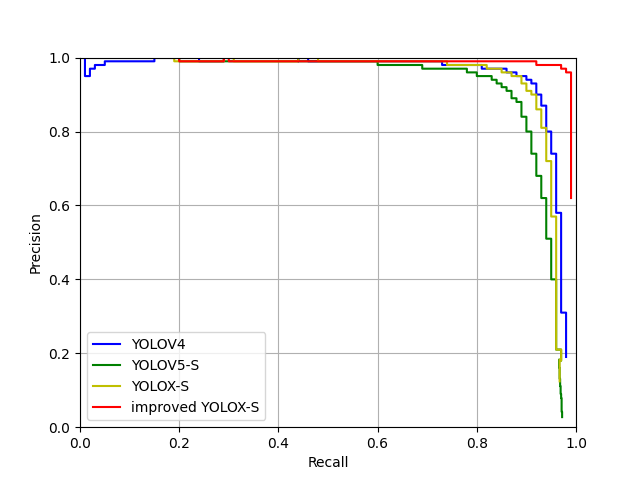


Fig. 5. The RP curves of different detector.

*-Table: The addition of FPS in Table I for real-time comparison.*

*Our response:*

Thanks to Reviewer#2 for pointing out the inadequacies in the table. We have added the related indicators of the performance of different detectors in **Table 2** on page 37.

**Table 2** Object detection results comparison.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Methods | F1 | Recall | Precision | AP50 | AP50:95 | FPS |
| YOLOV4 | 0. 88 | 84. 32% | 91. 49% | 90. 7% | 43. 7% | 44 |
| YOLOV5-S | 0. 88 | 87. 71% | 89. 28% | 91. 5% | 50. 9% | 94. 2 |
| YOLOX-S | 0. 91 | 89. 76% | 93. 24% | 93. 3% | 54. 4% | 79. 8 |
| Improved YOLOX-S | 0. 98 | 97. 95% | 97. 17% | 98. 4% | 70. 6% | 50. 7 |

*-Applications: For the micro-platform aspect mentioned in the paper, Is there an application for the actual farming environment?*

*Our response:*

As for the micro-platform not mentioned in the paper, we would like to thank reviewer #2 for the suggestions given in this comment. In the experimental process of this subject, improved YOLOX-S and Bytetrack are used to detect and track fish with abnormal behaviors. The reason why the lightest network structure in the standard version is selected for improvement and application is to ensure accuracy and achieve real-time monitoring. In the future practical application, it can be completely deployed in the micro-platform for unmanned monitoring.

*- Others: The results are best displayed using video, otherwise the tracking result is hard to observe .*

*Our response:*

We deeply agree with your comments, so we put the relevant videos shown in this article in the **Supplementary information**.