Response to Reviewer 3 Comments

Thank you for your comments concerning our manuscript. Those comments are all valuable for revising and improving our paper, as well as the important guiding significance to our researches. The font colour of the modified parts is red in the revised manuscript.

Point 1: The authors have proposed Underwater-YCC target detection algorithm so it should be described that what does YCC represent.

Response 1: Thanks for your careful reading and insightful suggestions. According to your suggestions,we have add CBAM and Conv2Former to YOLOv7 model, so we call it YCC(YOLOv7 with CBAM and Conv2Former, YCC). In addtion, we have given explanation in the text.

Point 2: In line 21, authors have compared their result accuracy with 'conventional target detection algorithms', which of the conventional algorithms are they referring to here?
Response 2: Thanks for your careful reading and insightful suggestions. According to your suggestions, the expressions in the abstract have been revised. In the paper, we compared Underwater-YCC with some mainstream models, like Faster-RCNN, YOLOv3，YOLOv5s，YOLOv6，YOLOv7-Tiny and YOLOv7, the experiment results show in table 5. In addition, similar expressions in the full text have been checked and revised.

Table5. Compare with classical target detection algorithms

Model	Precision	Recall	mAP	F1 Score	FPS
Faster-RCNN	38.3%	55.26	62.1%	45.24	16
YOLOv3	76.6%	64.3%	80.9%	69.91	
YOLOv5s	83.34%	78.96%	83.88%	81.09	46.51
YOLOv6	82.1%	63.6%	82.09%	71.67	
YOLOv7-Tiny	81.44%%	79.14%	84.21%	80.27	32.89
YOLOv7	84.21%	80.97%	85.67%	82.55	26.42
Underwater-YCC	84.64%	81.38%	87.16%	82.97	21.17

Point 3: In lines 102-103, authors state "Underwater-YCC has higher detection accuracy compared with the classical algorithm"-what do they mean by classical algorithm here? If it is basic YOLOv7 then mention it.

Response 3: Thanks for your careful reading and insightful suggestions. According to your suggestions, we have revised the expressions of classical algorithm. In addition, similar expressions in the full text have been checked and revised.

Point 4: Explain the different maps shown in section (a) of figure 1.

Response 4: Thanks for your careful reading and insightful suggestions. According to your suggestions, we have added the description to figure 1(a). Figure 1(a) shows the amount of data for each category, the size and number of bounding boxes, the location of the sample centroids, and the aspect ratio of the target occupying the entire image,respectively.

Point 5: In lines 229 and 237, authors should specify clearly that what does Mc(F) and Ms(F) denote.

Response 5: Thanks for your careful reading and insightful suggestions. According to your suggestions, we have revised the expressions in the paper. Here, parameters M_c (F) and M_s (F) represent the feature maps of the channel attention mechanism and the spatial attention mechanism respectively.

Point 6: In line 252-253, authors state "we will focus on the remaining three sections" -the use of the term 'remaining' doesn't add up here.

Response 6: Thanks for your careful reading and insightful suggestions. According to your suggestions, we have revised the remaining three parts as:" we will focus on Backbone, Neck, and Head".

Point 7: In line 302, authors may wish to state that what does 'F' stands for in ELAN-F structure.

Response 7: Thanks for your careful reading and insightful suggestions. Comparing Figure 9 and Figure 12, we can see that the ELAN-F structure is similar to the ELAN structure in the Backbone, but the ELAN-F structure is a sum of full each output section, in order to distinguish it from model ELAN, we call it ELAN-F structure.

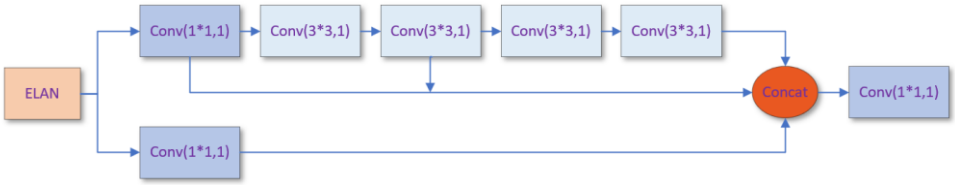


Figure 9. The architecture of ELAN

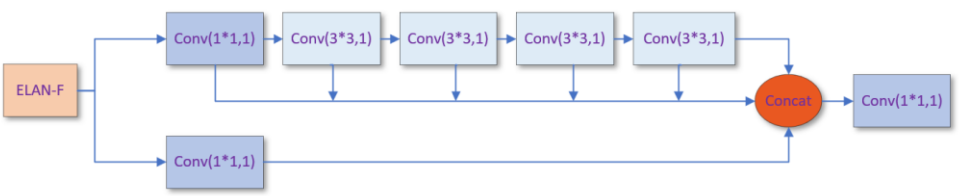


Figure 12. The architecture of ELAN-F.

Point 8: In line 375, rather than using the term 'calculate it in a specific way', authors should explain it thoroughly.

Response 8: Thanks for your careful reading and insightful suggestions. According to your suggestions, we have amended this inappropriate expression. In addition, we have checked the full text carefully to avoid similar wrong expression.

Point 9: In line 389, eq (6), what do the terms 'xgt' and 'ygt'denote?
Response 9: Thanks for your careful reading and insightful suggestions. According to your suggestions, we have explained the parameters in the equation (6).

An anchor box is represented by $\vec{B} = [x\ y\ w\ h]$, where the value represents the center coordinates and size of the corresponding bounding box, and $\overrightarrow{B_{gt}} = [x_{gt}\ y_{gt}\ w_{gt}\ h_{gt}]$ refers to the corresponding value of the target box. In the above equation.……

Point 10: In section 4, along with experimental environment, authors should also include experimental configurations such as batch size, epochs etc.
Response 10: Thanks for your careful reading and insightful suggestions. According to your suggestions,We have added the experimental configurations parameters in the table 1.

Table 1. Experimental environment and parameters

Configuration	Parameter
CPU	Intel(R) Core(TM) i9-10920X
GPU	NVIDIA GeForce RTX 3090
Operating system	Windows10
Frame	Pytorch1.7
CUDA	11.7
Batch Size	16
Epochs	300
Image Size	640*640

Point 11: The writing in this paper is difficult to follow due to poor grammar and sentence structure. There are many run-on sentences, incomplete thoughts, and unclear explanations throughout the paper. The few examples are:
Line 111 : "…ability of the model. In this paper…" should be single sentence as "……ability of the model, in this paper…".
Line 132 : "…training data, and lack…" should be "…training data and lack…".
Line 156 : "models with weak this ability" should be "models with this weak ability".

Line 186: “enchance” should be “enhance”.
Line 191: Sentence should not start from ‘and’.
Line 313: “…correspond to the a priori box” should be “…correspond to a priori box”.
Please revise the writing and proofread for other similar errors.

Response 11: Thanks for your careful reading and insightful suggestions. In the revised version, we have carefully checked and revised the whole text, all detectable typos and grammatical mistakes have corrected.