IMPORTANT NOTE (DO NOT DELETE)
The rebuttal to the reviews **must** be restricted to:

- answering the specific "pressing" questions raised by the reviewers in slot 4 of the reviews; and
- pointing out factual errors in the reviews.

Furthermore, it **must be one page-only**.
The goal of the rebuttal is not to enter a dialogue with the reviewers. Therefore, it is **not allowed to include new results in the rebuttal, provide links to such results, or describe how you might rewrite your paper to address concerns of the reviewers**. It is the current version of your paper that is under review, not possible future versions.
The program committee will be instructed to ignore rebuttals violating these principles.

Rebuttal

Thank you for all the reviewers and our answers as follows:

Reviewer #1:

Q1: Better to include the results of Baseline+RCG.

A1: Thanks for the suggestion. As you said, adding this experiment can clearly prove the effectiveness of the RCG module. We will present the results in the revised paper.

Q2: Whether multi-level information from E-blocks and multi-level features obtained in ALGM is redundant?

A2: Thank you. They are not redundant. The multi-level information extracted from E-blocks is insufficient and cannot adapt to polyps of different scales. While the function of ALGM is to enhance feature representation and get global and local multi-level features. In a word, the purpose and benefit of these two multi-level are different.

Reviewer #2:

Q1: Please offer some illustrations for how to address the false negative prediction.

Q2: There is no illustration of removing background noise.

Q3: Provide the visualization mentioned in Sec.3.3.

Q4: Should visualize feature f and the feature after MLP?

Q5: Should apply the same setting to generate a feature map for better understanding?

A(1,2,3,4,5): Thanks for the time and insightful feedback. We apologize for causing this confusion due to page limitations. The results of these visualizations are very significant and contribute to supporting our core ideas. According to your suggestion, we will add these intuitive visualizations in the final paper.

Q6: Why apply MLP rather than attention or a convolutional layer to generate a boundary location information?

A6: In our method, we mainly consider how to provide accurate pixel-level attention for the fused features extracted from reverse and boundary features. Due to the channel-wise attention lacking pixel-wise information[1][1]. Based on this, we use MLP to extract representative features, and then produce a boundary location information weight at pixel-level to guide the fused features learning. Particularly, the convolution layer does not meet the requirements in this process.

Q7: The ETIS, ClinicDB datasets should also be involved in the experimental section.

A7: Thanks. We agree with the reviewer that more datasets can improve the reliability of generalization capability. Experimental results will be fully presented in the final version.

Reviewer #3:

Q1: The main difference between RCG and reverse attention module? What advantages of RCG?

A1: Thank you very much. The key difference between the two modules is reverse attention module only applies attention to the regional background and introduces unwanted background noise, while RCG adds the edge information of the low-level to focus on unpredicted targets within the contour. Also, RCG compensates for edge detail loss in feature sampling needed for polyp segmentation.

Q2: Should you validate the effectiveness of the CBAM without using the edge features?

A2: Thanks for pointing out the study that we have missed. The CBAM consists of two attention operations. It can suppress irrelevant background noise and capture polyp detail information from the low-level feature map. We will include the suggested study in the final version.

Q3: Some experimental settings are not clear.

A3: In learning ability, we refer to [2][2] to divide dataset with 60% as training set, 20% as validation set and 20% as test set. In generalization capability, the Kvasir-SEG and EndoScene-CVC612 are split into 90% for training. Please see Section 4.3. For other model parameters, some are provided in Section 1 of supplementary material, and we will give a more detailed list in the attachment of the final version.

Q4: S-measure and E-measure metrics need to be added.

A4: Thanks for the comments. We refer to [2][2] to adopt eight widely-used metrics. We agree with the reviewer that S-measure and E-measure metrics are necessary to compare and we will add them in Table 2 in the final version.

Q5: In the ablation study, when without using HPPF, how to produce the final prediction map?

A5: Thanks. It just uses the output of D-Block1 to produce the final prediction map, please see the Figure S2 of supplementary material.

Q6: The used symbols are particularly messy.

A6: Thank you very much. We will carefully revise them in the final version.

Reviewer #4:

Q1: It would be interesting to consider the problem of uncertainty related to the raters disagreement.

A1: This is a very good suggestion. It is valuable to the problem of uncertainty related to the raters disagreement and we will provide it in the final version.

Q2: If many raters were annotated datasets, how your method takes into account the divergence or the disagreement?

A2: Thanks. When the polyp is labeled by multiple clinicians. For better consistency, the average geometric distance between annotations is computed. According to a given pixel error margin, retain the annotations of 2D Hausdorff distance is smaller than or equal to margin. And based on that, the final label is determined by voting, further convert the labels to grayscale images and normalize the pixels to 0 and 1.

[1] [1] Li et al., Pyramid Attention Network for Semantic Segmentation[J]. CVPR 2018

[2] [2] Zhang et al., Adaptive context selection for polyp segmentation[C]. MICCAI 2020