

Dear Reviewer and Editor,

Thanks again for the comments on our manuscript “Enhanced segmentation of gastrointestinal polyps from capsule endoscopic images with artifacts using ensemble learning” (Manuscript NO: 78498). We have read and considered all of the comments and revised the manuscript based on the reviewer’s recommendations. Our point-by-point responses to each reviewer’s comments are appended below. The parts of our revision have been highlighted in the article.

Thanks again for your help to our paper processing. We are looking forward to hearing from you soon.

Best regards,

Hong Wang (on behalf of all authors)

Department of Gastroenterology and Hepatology,

Guangzhou First People’s Hospital,

School of Medicine, South China University of Technology,

E-mail: wong.hong@163.com

No.1 Panfu Road, Yuexiu District

Guangzhou 510180, P.R.CHINA

Point-by-point Responses to Reviewer Comments

Reviewer #1:

Specific comments:

1. Major points Discussion is too short. Please discuss the strengths and limitations more in detail.

RE: Thanks for your suggestion, and we have revised the Discussion carefully. The parts of our revision have been highlighted in the Discussion.

2. Introduction: OK, but please explain why the authors focused on GI polyps only, not on a wide range of GI diseases.

RE: Thank you very much, and the parts of our revision have been highlighted in the introduction. As shown below.

Colorectal cancer (CRC) is the second leading cause of death in the USA^[1]. In China, an estimated 1,101,653 new cancer cases and 709,529 cancer deaths from gastric cancer and CRC will occur in 2022, placing China first worldwide because of its large population^[1]. Although other gastrointestinal lesions, such as erosions and ulcers, can also develop into cancers, most gastrointestinal cancers arise from precancerous polyps, which are the most common lesions found on endoscopy^[2]. Therefore, early detection and removal of gastrointestinal polyps under endoscopy are critical for preventing gastrointestinal cancers^[3-6].

Due to the variety of polyp shapes, some smaller and flat polyps are easily missed, which delays diagnosis and treatment. Therefore, many scholars have focused on the automatic detection of polyps." This is why we only focused on gastrointestinal polyps in this article. And we will focus on other gastrointestinal diseases in follow-up experiments.

3. Materials and Methods: This paper is the first combined study of semantic segmentation and ensemble learning to analyze CE images with artifacts. Thus, please describe ensemble learning used in this study more in detail. Custom-

built model?

RE: Thanks for your suggestion. We have added a description of the ensemble learning method we use in the article. The method we use comes from an ICLR 2017 paper called snapshot ensemble learning, which is an ensemble learning method proposed for deep learning methods. The main idea is to use an SGD optimizer to make the model converge to a flat local optimum, and then restart the learning rate several times to let the model find multiple such flat local optimums. The inference of the model is performed by the vote ensemble method. We have added details of our use of this method in the revised version of the paper, including information on the parameters used, the schedule of the learning rate, and the steps where local optima are obtained.

4. Discussion: Please review briefly the limitations of previous A.I. studies without ensemble learning and discuss future direction for A.I. application for CE imaging.

RE: Thank you very much, and the parts of our revision have been highlighted in the Discussion.

Limitations:

In the field of machine learning, ensemble learning is usually used to reduce model variance. Variance reduction enhances the capability of the model against image noise. In other words, when the prediction of the model varies greatly but is relatively accurate (with low bias) in general, we consider using ensemble learning methods to make the model less sensitive to training details to obtain relatively stable prediction results. Thus, the disadvantage of studies without ensemble learning methods is the large variance of the model. Conversely, the usefulness of ensemble learning methods is limited when the variance of the model is already low. Müller's work shows that ensemble learning methods improve the robustness of medical image classification in a credible manner. Therefore, models without ensemble

learning methods do not perform well on images with artifacts in the real world.

Future direction:

We believe that the direction of feature AI for CE imaging research lies in making existing computer models better serve clinical diagnosis in a practical sense rather than letting these methods stay in the laboratory. CE is commonly used to examine digestive diseases. In addition to polyps, many digestive diseases can be detected using CE. Thus, AI for CE imaging can be considered to enrich the diagnosis, localization, and grading of more forms of the disease, such as ulcers and erosions, to assist doctors in more refined disease research and diagnosis. Additionally, from the perspective of deep learning, better model performance often relies on more model parameters and computational resources. In the future, we will validate the ensemble learning method in clinical practice to demonstrate that it can improve the detection rate of polyps in CE in the clinic and evaluate the potential of this method for other types of medical images or lesions.

5. Please discuss about the false negative case.

RE: Thank you very much, and the parts of our revision have been highlighted in the Discussion. As shown below.

From the experimental results, although the ensemble learning approach improves the performance of the segmentation model on the dataset with artifacts, there are still false-positive and false-negative cases. On the one hand, the main reason for false-negative cases is that the model confuses normal-color polyps with normal gastrointestinal folds or confuses abnormal-color polyps with artifacts, such as yellow bubbles. However, the main cause of false-positive cases was that some artifacts or normal folds had a high similarity with polyps in the image, which led the model to misidentify them as polyps. Overall, the main reason for segmentation errors is that the color and texture are highly confusing, and we will further attempt

to improve the ability of the model to distinguish polyps, normal tissues, and artifacts in a subsequent study.

6. Please discuss “cost” and “complexity” of the system used in this study, because the A.I. calculation algorithm require specialized software packages, leading to increased costs, and the workflow of image acquisition, segmentation, feature extraction, exploratory analysis and modeling makes the system extremely complex.

RE: Thank you for your suggestion, and the parts of our revision have been highlighted in the Discussion.

We are sorry for not able to describe the cost of using ensemble method and not using this method in the original article in detail. It should be clear that ensemble learning is a process of packaging multiple weak models together to form a strong model and does not need for more input information. Therefore, in the whole workflow process including image acquisition, exploratory analysis, feature extraction, modeling, etc. , the only additional cost lies in the computational resources consumed by using our method in the modeling process, including computational power and time. The method does not add additional steps to the workflow. Generally, the cost of our method is the additional computational resources needed to train the model, but not the additional steps in workflow.

Reviewer#2

Specific comments:

1. Please describe in detail how many images were used for training, validation, and test sets, respectively, in your dataset and public dataset.

RE: Thanks for your suggestion. We randomly divided the experimental and public data into training (195 images), validation (41 images), and testing (41 images) sets.

2. Are polyps present in all images used?"

RE: Thank you very much, and all images used in this study contained at least one polyp class, including the standard datasets. To ensure the accuracy and rigor of the data annotation, the image data were obtained by an experienced gastroenterologist who watched the video recordings, extracted the frames where the polyps were captured through ES Navi, and annotated the pixel points of the polyp lesions using Labelme. Next, the annotated polyp profiles were carefully reviewed by two other experienced gastroenterologists.

3. If the ultimate purpose of the study is to recognize the polyp of CE, it is thought that the test should be performed with the CE image.

RE: Thank you for your valuable comments. Our ultimate goal is not only to identify capsule endoscopy (CE) polyps, but also to show that ensemble learning can perform better on the dataset with artifacts. Therefore, we performed the same experiment on GZ_Capcam, CVC_Clinic and CVC_Colon datasets. GZ_Capcam is capsule endoscopy data set. CVC_Clinic, and CVC_Colon are colonoscopy data sets. The comparison of experimental results on the three data sets shows that the ensemble learning method can improve the performance of the model on the artifact data set (GZ_Capcam, capsule endoscopy data set). And the degree of improvement is greater than the degree of improvement on clear data sets (CVC_Clinic and CVC_Colon, colon endoscopy dataset). As your comment suggests, we used CE data for our tests. The experiment using the colonoscopy dataset is just to more rigorously discuss the validity of the method we used.

4. Please describe the clinical significance of the increase in polyp detection segmentation accuracy derived from the results.

RE: Thank you very much, and the parts of our revision have been highlighted in the Discussion. As shown below.

Previous studies have extensively analyzed and concluded that integrated learning methods improve the robustness of medical image classification in a credible manner^[32]. By improving the segmentation performance of the model, we can separate polyps more accurately from surrounding tissues, which can improve the detection probability of polyps and aid in monitoring the size of polyps in patients with unresectable polyps^[33-35]. Semantic segmentation provides pixel-level classification and clearer polyp boundaries, which are also crucial in surgical procedures or radiofrequency ablation and is expected to be used for real-time detection of polyp boundaries in surgical resection under gastroenteroscopy to assist polyp resection^[33, 36].

5. When the ensemble model is applied, it is estimated that the time required for analysis is longer than that of a single model. Please describe if the ensemble model has any disadvantages other than increasing the recognition rate.

RE: Thank you very much, the parts of our revision have been highlighted in the Discussion. As shown below.

The use of the ensemble learning approach mentioned in this paper does not imply any additional workload or workflow reordering. The only additional cost associated with the method is the computational resources. Additionally, from the perspective of deep learning, better model performance often relies on more model parameters and computational resources. Methods that already use deep learning models can easily apply ensemble learning methods to improve model performance without the need for additional workflow tuning. It is worth mentioning that, although the ensemble learning approach can improve the robustness of CAD image analysis models, misuse may lead to a less-than-expected improvement in the model's effectiveness, mainly because the essence of ensemble learning is to reduce model variance, and when the variance of a single model is already very low, the improvement brought by ensemble learning may be

very limited. In clinical practice, video frames can be completely infested with artifacts, making the content of the image simply unrecognizable. Therefore, the appearance of these frames is inevitable in clinical practice. In the present study, we confirmed the authenticity of polyps in pictures with artifacts by using more images, videos, and other inspection methods. Thus, we solved the dilemma of applying AI to these medical images. However, our study has some limitations. For example, the images were insufficient and did not involve lesions other than polyps.