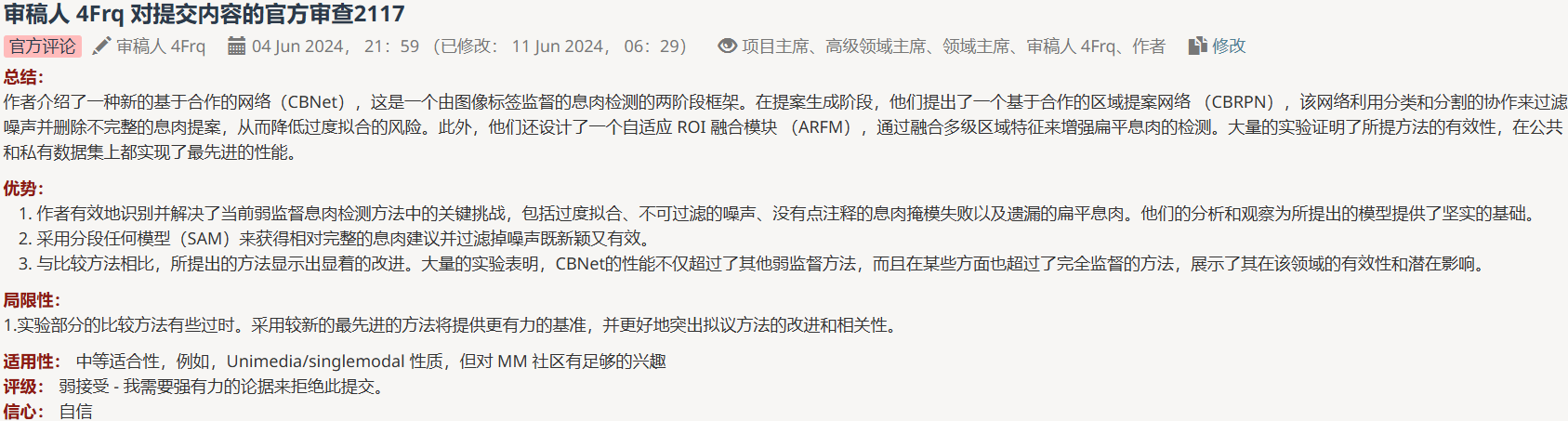
# Review #1



1.The comparison approaches in the experimental section are somewhat outdated. Incorporating more recent state-of-the-art approaches would provide a stronger benchmark and better highlight the improvements and relevance of the proposed method.

# Review #2



1.The writing quality of this paper is not particularly good. Some sentences need to be clarify:

In lines 111-113, the authors point out that the MIL-based and SAM-aided approaches have complementary advantages and disadvantages. However, the authors do not mention about the disadvantages.

In line 135-138, the authors mention about "the former" and "the latter". "The former" and "the latter" mean what?

What is SSW. In the related work section, authors only mention about the Selective Search (SS).

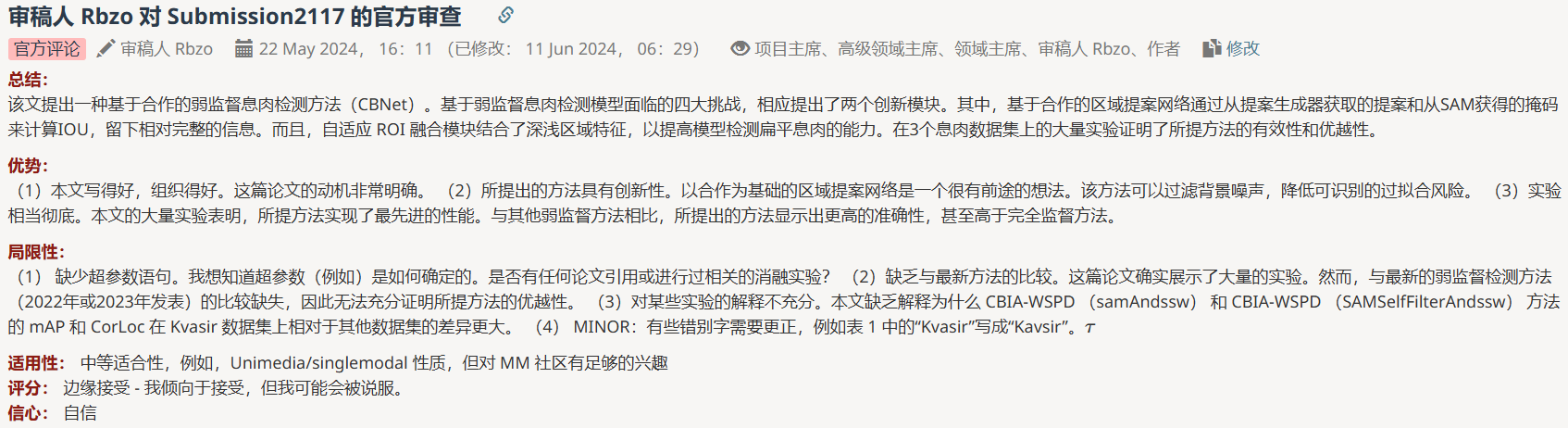
In line 370-374, this sentence could be rewritten for better clarity. the CBRPN takes the image I and feature F\_1 - F\_5 as inputs. The image I is used to generate the B\_SSW and the feature F\_1 - F\_5 are used to produce the coordinates.

The authors should mention the WSDDN in the related work section. Since the MIDN is built upon the WSDDN.

The caption of Table 1 need to be rewrite. Since the best indicator of CBNet is red font and the best indicator of weak supervision is blue font. Besides, What is CBIA-WSPD in Table 1?

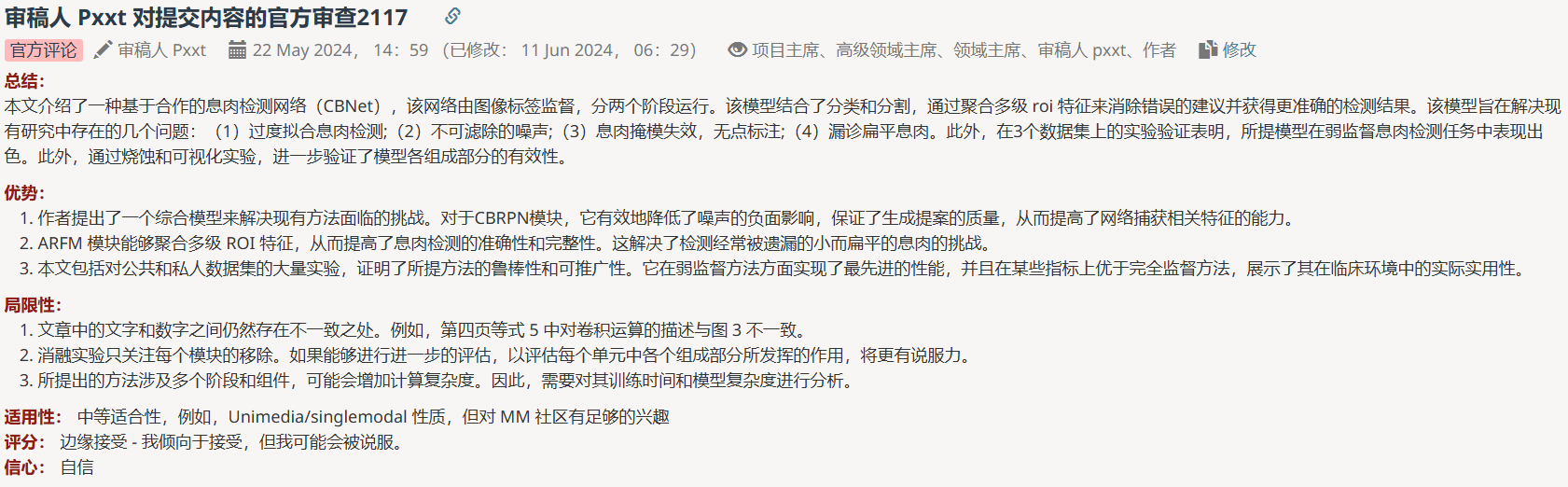
2.In Table 2, why SAM and SAM(filter) do not have results?

# Review #3



1. Missing statement of hyperparameters. I would like to know how the hyperparameters (e.g. ) are determined. Are there any papers referenced or relevant ablation experiments performed?
2. Lack of comparison with the latest methods. This paper does show extensive experiments. However, the comparison with the latest weakly supervised detection method (published in 2022 or 2023) is missing, and thus the superiority of the proposed method cannot be fully proved.
3. Inadequate explanation of some experiments. The paper lacks the explanation of why the mAP and CorLoc of the CBIA-WSPD (samAndssw) and CBIA-WSPD (samSelfFilterAndssw) methods differ more on the Kvasir dataset relative to the other datasets. (4) MINOR: Some typos need to be corrected, i.e. 'Kvasir' is written as 'Kavsir' in Table 1. τ

# Review #4



1. There is still inconsistency between the text and the figures in the article. For instance, the description of the convolution operation in Equation 5 on page four is inconsistent with Fig. 3
2. The ablation experiments only focused on the removal of each module. It would be more persuasive if further evaluation could be conducted to assess the roles played by individual components within each module.
3. The proposed method involves multiple stages and components, which may increase the computational complexity. Therefore, an analysis of its training time and model complexity is expected.
4. 我们的贡献CBRPN({ssw},{sam})和ARFM({Conv\_P & GAP},{Agg & Fusion})在CBNet的role如图（a）所示，因此CBRPN必须保留其中的一个组件，各组件的评估结果请看表。根据您的建议我们进一步对ARFM的组件进行了评估，请看表。当只有组件1时map为，当只有组件2时map为，同时存在时map为（proposals来自SSW&SAM on clinicdb）。

# Summary

Review #2,3 ,4:Some error need to be corrected. Add WSDDN to related work.

Review #1,3 : Lack of comparison with the latest methods.

Review 2

Q1: Explain the disadvantages of MIL-based and SAM-aided approaches, the mean of ‘former’ and ‘latter’(line 135-138), the mean of SSW and CBIA-WAPD.

A1: The disadvantage of MIL-based is overly focus on the most discriminatory local regions, resulting in incomplete prediction, while the disadvantage of SAM-aided is that the work requires the most discriminatory point annotations;‘former’is a proposal generator (SSW) ,’latter’is SAM ; SSW is Selective Search Windows, CBIA-WSPDis an editing error that has been corrected to CBNet.

Q 2:In Table 2, why SAM and SAM(filter) do not have results?

A2:

Review 3:

Q1:How are hyperparameters（τ） determined？

A1:

Q2:why the mAP and CorLoc of CBNet more different on the Kvasir.

A2:

Review 4:

Q1：Assess the roles played by individual components within each module.

A1：我们的贡献CBRPN({ssw},{sam})和ARFM({Conv\_P & GAP},{Agg & Fusion})在CBNet的roles如图（a）所示，因此CBRPN必须保留其中的一个组件，各组件的评估结果请看表。根据您的建议我们进一步对ARFM中的组件进行了评估，组件2不可单独评估因为输出是MIDN（网络的必有模块）的输入，组件1的评估结果请看表。当只有组件1时map为，当只有组件2时map为，同时存在时map为（proposals来自SSW&SAM on clinicdb）

Q2: Analyse the training time and model complexity .

A2: