

# View Reviews

## Paper ID

1662

## Paper Title

Solving Ambiguous Boundary: Leveraging CORAL-Correlation Aligning Semi-Supervised Medical Image Segmentation

## Reviewer #1

---

### Questions

#### 4. Please describe the contribution of the paper (a few lines).

In this paper, the authors proposed a CORAL-Correlation Constancy Framework(CORF), a semi-supervised learning schemes based on cross-learning. Previous methods face challenges in the issues of ambiguity in segmentation edges and accumulation of errors in pseudo-labels derived from unlabeled images. To address above problems, the proposed CORF dynamically adjust to achieve alignment between the distribution of labeled and unlabeled data to better capture complex organ structures. Besides, the incorporated Dynamic Feature Pool(DFP) discards features with low segmentation confidence and incorrect predictions, and the consistency strategy for model regularization only retains consistent features from the dual encoder that are significant, thereby improving model performance and reliability. The quantitative and qualitative experimental results show the effectiveness of the proposed method.

#### 5. Please list the main strengths of the paper; you should write about a novel formulation, an original way to use data, demonstration of clinical feasibility, a novel application, a particularly strong evaluation, or anything else that is a strong aspect of this work. Please provide details, for instance, if a method is novel, explain what aspect is novel and why this is interesting.

- a) The writing of this paper is easy to understand and there is few spelling or grammatical error.
- b) The quantitative and qualitative experimental results show the effectiveness of the proposed method, and the article is SOTA in comparative experiments.
- c) Different the previous methods only focus on pixel level, CORF introduces second-order statistical features (co-variance) of labeled and unlabeled feature distributions resulting in more advantages about extracting global and local features compared to cosine similarity.
- d) The method notices the misrepresentation of unlabeled data. The Dynamic Feature Pool (DFP) mechanism employs a confidence-based filtering and consistency strategy, mitigating the negative impact of ineligible features derived from unlabeled data.

e)The method tries to describe the network architecture as well as the objective functions thoroughly.

**6. Please list the main weaknesses of the paper. Please provide details, for instance, if you state that a formulation, way of using data, demonstration of clinical feasibility, or application is not novel, then you must provide specific references to prior work**

a)In "Introduction" section, the review of existing semi-supervised medical image segmentation methods [1-5] is insufficient. Semi-supervised medical image segmentation is a area of high interest. However the majority of references in the paper are conducted on natural condition, and there is no in-depth discussion of development in semi-supervised methods of medical area. Besides, the authors should clearly state the differences between their method against these exiting.

b)In method section, there is incorrect use of symbols about CORR, specifically, writing CORRmp as CORRvp and CORRap as CORRmp in the same paragraph, besides the loss  $L_c$  is not described in as much detail as  $L_s$  while they both come from the CPS framework.

c)In "Experiments and Results" section, when it comes to ablation study, the authors claims that "Given that the DFP is a pivotal component for the CCM to filter and recommend the high-quality anchor points, the study is structured to compare the CORAL-Correlation Constancy Computation to the traditional cosine similarity method [13,22] in terms of similarity computation", therefore no separate ablation experiments were performed for DFP which fails to prove the validity.

d)Although the authors give the clear motivations for the study, but doesn't explain how CORP solves the problem of ambiguous boundaries in detail. In a more fine-grained way, why "second-order statistics information (covariance) of labeled and unlabeled feature distributions has more advantages in extracting global and local features compared to cosine similarity, which focuses on pixel-level information, especially in areas with ambiguous boundaries." As the ambiguous boundaries is the key of the paper title, it deserves more elaboration. If the theory is derived from other references, specific examples should be given.

[1]Wen, Lu, et al. "DCL-Net: Dual Contrastive Learning Network for Semi-Supervised Multi-Organ Segmentation." ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2024.

[2]Zhang, Zhenxi, et al. "Self-aware and cross-sample prototypical learning for semi-supervised medical image segmentation." International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature Switzerland, 2023.

[3]Basak, Hritam, and Zhaozheng Yin. "Pseudo-label guided contrastive learning for semi-supervised medical image segmentation." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023.

[4]Tang, Cheng, et al. "Semi-supervised medical image segmentation via hard positives oriented contrastive learning." Pattern Recognition 146 (2024): 110020.

**7. Please rate the clarity and organization of this paper**

Satisfactory

**8. Please comment on the reproducibility of the paper. Please be aware that providing code and data is a plus, but not a requirement for acceptance.**

The authors claimed to release the source code and/or dataset upon acceptance of the submission.

**10. Please provide detailed and constructive comments for the authors. Please also refer to our Reviewer's guide on what makes a good review. Pay specific attention to the different assessment criteria for the different paper categories (MIC, CAI, Clinical Translation of Methodology, Health Equity):**

<https://conferences.miccai.org/2024/en/REVIEWER-GUIDELINES.html>

a) In "Introduction" section, more semi-supervised segmentation methods need to be reviewed, especially semi-supervised medical image segmentation methods. The representative ones should be specifically summarized to better magnify the difference and superiority of the proposed method.

b) Correct the formula error, and elaborate the method more detailed and fairer.

c) DFP ablation study should be added.

d) More elaboration about how the proposed module solve the problem comes from the motivation. If the theory is derived from other references, specific examples should be given.

**11. Rate the paper on a scale of 1-6, 6 being the strongest (6-4: accept; 3-1: reject). Please use the entire range of the distribution. Spreading the score helps create a distribution for decision-making (visible to authors).**

4. Weak Accept — could be accepted, dependent on rebuttal

**12. Please justify your recommendation. What were the major factors that led you to your overall score for this paper?**

a) The lack of ablation study about DFP leads to questions about the authenticity of the experimental data and the validity of the module. The superiority of the proposed module cannot be proved only in theory from the author's perspective.

b) Some important issues in writing, which mainly consist of incorrect formulas, insufficient review and explanation about how the ambiguous boundaries problem is solved.

**14. Reviewer confidence: In view of your answers above and your overall experience, how would you rate your confidence in your review?**

Confident but not absolutely certain (3)

**19. [Post-rebuttal only] After reading the author's rebuttal, state your overall opinion of the paper.**

4. Weak Accept — could be accepted, dependent on rebuttal

**20. [Post-rebuttal only] Please justify your decision**

Considering the submitted rebuttal and other papers in my stack, i would like to insist my score.

## Reviewer #3

## Questions

**4. Please describe the contribution of the paper (a few lines).**

This paper tackles the semi-supervised segmentation problem, where a CORAL-correlation constancy framework is proposed. CORAL is a correlation matrix in UDA task (2016). Distinct from previous methods, CORF uses second-order statistic features to capture global structural information to minimize the shift between labeled and unlabeled data. The method is validated on the LA benchmark only.

**5. Please list the main strengths of the paper; you should write about a novel formulation, an original way to use data, demonstration of clinical feasibility, a novel application, a particularly strong evaluation, or anything else that is a strong aspect of this work. Please provide details, for instance, if a method is novel, explain what aspect is novel and why this is interesting.**

The paper is well-written and easy to follow.

Code will be available.

**6. Please list the main weaknesses of the paper. Please provide details, for instance, if you state that a formulation, way of using data, demonstration of clinical feasibility, or application is not novel, then you must provide specific references to prior work**

The writing should be improved. The logic in the introduction seems redundant and confusing, especially in the context of method introduction. You don't need to separate these contexts into different paragraphs.

Authors aims at the ambiguity in segmentation edges (boundaries), yet, no metric or other theoretical analysis can support this. I don't think the Figure 1 can be a good support for this claim. It is just a typical presentation of segmentation results.

Page 4: The loss symbols should be consistent, using  $\mathcal{L}$ . What are  $L_s$  and  $L_c$  in CPS?  $L_s$  means supervised loss on the labeled data, right?  $L_c$  is the cross-supervision loss? The writing is not clear.

What are the supervised losses for the compared methods? All CE loss or the combination of CE and Dice losses?

One biggest concern is that the method is only validated on the left atrium dataset, yet claiming SSL for medical image segmentation. Since the LA benchmark is well-solved now and please add other challenging datasets for comparison if claiming medical image segmentation. Besides, the claim in the abstract saying that "this design can ... complex organ structures", the current experiment is hard to demonstrate this because the task itself is not so challenging. Authors should apply the methods to more complicated data, especially for multi-class segmentation. So far, some of the contexts

and motivation are overclaimed without strong support.

How about the computation cost for the method?

The authors claim distribution alignment yet no experiment or visualization can support it.

What is the temperature value?

What is the insight on the performance even surpasses the upperbound (VNet)?

**7. Please rate the clarity and organization of this paper**

Good

**8. Please comment on the reproducibility of the paper. Please be aware that providing code and data is a plus, but not a requirement for acceptance.**

The submission has provided an anonymized link to the source code, dataset, or any other dependencies.

**10. Please provide detailed and constructive comments for the authors. Please also refer to our Reviewer's guide on what makes a good review. Pay specific attention to the different assessment criteria for the different paper categories (MIC, CAI, Clinical Translation of Methodology, Health Equity):**

<https://conferences.miccai.org/2024/en/REVIEWER-GUIDELINES.html>

See weakness.

**11. Rate the paper on a scale of 1-6, 6 being the strongest (6-4: accept; 3-1: reject). Please use the entire range of the distribution. Spreading the score helps create a distribution for decision-making (visible to authors).**

2. Reject — should be rejected, independent of rebuttal

**12. Please justify your recommendation. What were the major factors that led you to your overall score for this paper?**

Limited experimental design.

The claim is not well supported.

**14. Reviewer confidence: In view of your answers above and your overall experience, how would you rate your confidence in your review?**

Very confident (4)

**19. [Post-rebuttal only] After reading the author's rebuttal, state your overall opinion of the paper.**

2. Reject — should be rejected, independent of rebuttal

**20. [Post-rebuttal only] Please justify your decision**

The authors didn't well address my concerns. Besides, only the simple LA task is not sufficient.

## Questions

### **4. Please describe the contribution of the paper (a few lines).**

This paper proposes a CORAL-Correlation Constancy Framework equipped with a Dynamic Feature Pool and a confidence-based filtering strategy to solve the problem of ambiguous boundary in semi-supervised medical image segmentation tasks. The experimental results on the Left Atrium dataset are impressive.

### **5. Please list the main strengths of the paper; you should write about a novel formulation, an original way to use data, demonstration of clinical feasibility, a novel application, a particularly strong evaluation, or anything else that is a strong aspect of this work. Please provide details, for instance, if a method is novel, explain what aspect is novel and why this is interesting.**

-The motivation of this paper is clear and reasonable.

-The performance of the proposed method is impressive, especially when the number of the labeled data is very small.

### **6. Please list the main weaknesses of the paper. Please provide details, for instance, if you state that a formulation, way of using data, demonstration of clinical feasibility, or application is not novel, then you must provide specific references to prior work**

-Limited novelty. The proposed Dynamic Feature Pool and the confidence-based filtering strategy seem to be the same as a previous work, i.e. CAML (Reference [2] in this paper). This greatly challenges the novelty and contribution of this work.

-Rationale of the method design. It's difficult to obtain the conclusion that the proposed method can solve the ambiguous boundary problem from the formulations. Moreover, the whole framework is very similar to CAML which only claims to be able to transfer the label prior knowledge from labeled data to unlabeled data. Moreover, pixels on the ambiguous boundary tend to have a low confidence, will this cause that the proposed Dynamic Feature Pool never selects ambiguous boundary pixels?

-Limited experiments. The proposed method is only validated on one dataset, i.e., the LA dataset whose shape is relatively fixed. It would be more interesting to see the effectiveness on a more challenging task, such as the tumor segmentation task.

### **7. Please rate the clarity and organization of this paper**

Good

### **8. Please comment on the reproducibility of the paper. Please be aware that providing code and data is a plus, but not a requirement for acceptance.**

The submission has provided an anonymized link to the source code, dataset, or any other dependencies.

### **10. Please provide detailed and constructive comments for the authors. Please also refer to our Reviewer's guide on what makes a good review. Pay specific attention to the different assessment criteria for the different paper categories (MIC, CAI, Clinical Translation of Methodology, Health Equity):**

<https://conferences.miccai.org/2024/en/REVIEWER-GUIDELINES.html>

-More clarification of how the proposed method can solve the ambiguous boundary problem theoretically. Also, it would be more helpful to show some visualizations of which pixels are often selected in the training process. Are they usually located on the boundaries? If no, it would be quite confusing about how the proposed method can solve the ambiguous boundary problem and what the really useful module in this proposed method is.

-More comparisons with related methods that aim to solve the ambiguous boundary problem as well, such as [1].

-More experiments on other datasets, especially more challenging tasks than the LA dataset.

-More description of CORAL and why it is selected in this work. It would also be better to provide a detailed description of the difference between reference [14] in the paper and the proposed method.

[1] Xu, Zhe, et al. "Ambiguity-selective consistency regularization for mean-teacher semi-supervised medical image segmentation." Medical Image Analysis 88 (2023): 102880.

**11. Rate the paper on a scale of 1-6, 6 being the strongest (6-4: accept; 3-1: reject). Please use the entire range of the distribution. Spreading the score helps create a distribution for decision-making (visible to authors).**

3. Weak Reject — could be rejected, dependent on rebuttal

**12. Please justify your recommendation. What were the major factors that led you to your overall score for this paper?**

Although the motivation of this work is reasonable and the performance is impressive, the method design is very similar to the previous work (reference [2] in the paper), especially the proposed Dynamic Feature Pool and the confidence-based filtering strategy, which seems to be the same as reference [2]. Also, how the proposed method can solve the ambiguous boundary problem is not clear.

**14. Reviewer confidence: In view of your answers above and your overall experience, how would you rate your confidence in your review?**

Confident but not absolutely certain (3)

**19. [Post-rebuttal only] After reading the author's rebuttal, state your overall opinion of the paper.**

2. Reject — should be rejected, independent of rebuttal

**20. [Post-rebuttal only] Please justify your decision**

The authors didn't well address my concerns on how the proposed method can solve the ambiguous boundary problem, which is also mentioned by all the other reviewers. In the rebuttal, they only show that the second-order statistics can capture the global structures and continuous features. However, its relationship to the ambiguity is unknown. Besides, the differences from CAML are not novel enough.