View Reviews

Paper ID
1565

Paper Title
ICNet: Intra-saliency Correlation Network for Co-Saliency Detection

Reviewer #1

Questions

1. Summary and contributions: Briefly summarize the paper and its contributions.
The co-saliency task is addressed in the paper. The authors consider the intra- and inter-image saliency together to improve the co-saliency results. The former is extracted with normalized masked average pooling and pre-computed single-image saliency maps. For the later, the correlation operator is applied to all images to capture the inter-image cues. Besides, to address the problem of the model that can't recognize objects with similar semantics, the self-feature scheme is proposed the category-independent signals to refine the co-saliency maps. In the experimental results, the proposed method outperforms the current state-of-the-art methods, especially on Cosal2015 dataset. Besides, the detailed ablation studies are conducted to validate the robustness of the proposed method.

2. Strengths: Describe the strengths of the work. Typical criteria include: soundness of the claims (theoretical grounding, empirical evaluation), significance and novelty of the contribution, and relevance to the NeurIPS community.
1. The proposed method is reasonable. The proposed rearranged self-correlation feature is novel and interesting, and it can contribute to the large performance gain.

2. The paper is clearly presented and well-organized.

3. The proposed method outperforms the current state-of-the-art methods, and the detailed ablation studies are conducted to validate the robustness of the proposed method.

3. Weaknesses: Explain the limitations of this work along the same axes as above.
1. Some related works missing
There are some recent related works, such as [Ref. 1~Ref.3], and it is better to cite these papers and have some discussion.

[Ref. 1] Zhang et al., "Adaptive Graph Convolutional Network with Attention Graph Clustering for Co-saliency Detection," CVPR'20
[Ref. 2] Fang et al., "Taking a Deeper Look at Co-Salient Object Detection," CVPR'20
[Ref .3] Tsai et al., "Deep Co-saliency Detection via Stacked Autoencoder-enabled Fusion and Self-trained CNNs," TMM'19

2. Overclaimed contribution
The proposed method contains many components, such as the combination of intra- and inter-image saliency, correlation fusion module, the normalized masked average pooling, rearranged self-correlation feature. However, the combination of intra- and inter-image saliency is done in [9], and the correlation fusion module is adopted in [9], too. Besides, the normalized masked average pooling has been proposed in [23]. These three should not be claimed or emphasized as the paper's contribution, but the authors should emphasize the rearranged self-correlation component. The proposed method is similar t to [9], which considers both intra- and inter-image saliency and the correlation module. The major differences are the normalized masked average pooling and rearranged self-correlation features, but the former has been proposed in [23]. Therefore, I think only considering rearranged self-correlation, the novelty is somewhat not enough to be accepted.

3. About the experimental results

3-1. The proposed method got a significant performance gain in Cosal 2015 dataset, but in the other two datasets, the only small performance gain is achieved. The authors should give a detailed discussion about it.

3-2. The model is trained with ten images from an image group, but the co-saliency maps could be generated for an arbitrary number of images. The training and inference seem not consistent. What are the results of the different images in an image group for training?

4. Correctness: Are the claims and method correct? Is the empirical methodology correct?
Yes

5. Clarity: Is the paper well written?
Yes

6. Relation to prior work: Is it clearly discussed how this work differs from previous contributions?
No. Please see Weaknesses

7. Reproducibility: Are there enough details to reproduce the major results of this work?
No

9. Please provide an "overall score" for this submission.
6: Marginally above the acceptance threshold.

10. Please provide a "confidence score" for your assessment of this submission.
5: You are absolutely certain about your assessment. You are very familiar with the related work.

11. Have the authors adequately addressed the broader impact of their work, including potential negative ethical and societal implications of their work?
Only partially, more discussion is needed.

Reviewer #2

Questions

1. Summary and contributions: Briefly summarize the paper and its contributions.
This work presents a Co-SOD method by utilizing the abundant information from off-the-shelf SOD method for extracting intra cues and deeply exploring inter cues in feature-level among an image group. The contributions mainly focus on the exploring intra and inter cues in a single image groups under the reference of SISM, and enhance this ability by RSCF.

2. Strengths: Describe the strengths of the work. Typical criteria include: soundness of the claims (theoretical grounding, empirical evaluation), significance and novelty of the contribution, and relevance to the NeurIPS community.
This paper skillfully explores intra and inter cues by several modules and achieves appealing performance which exceeds previous works a large margin in MAE metric. The proposed CFM module leverages SISMs as references, extracts intra cues in SIVs and explores inter cues (CSA maps) among an image group. The ablation studies and experimental results prove the effectiveness of the proposed modules.

3. Weaknesses: Explain the limitations of this work along the same axes as above.
1. The ablation studies base on Cosal2015 dataset and show convincing results. However, this work only obtains large improvement in this dataset while get tiny one in others, especially in terms of F-measure and S-measure. So, will this method also prove itself by performing ablation on other datasets? By the way, I cannot get your idea about the ablation setting for the CFM module, that is, how to concatenate a "group-level" vector with feature?
2. According to figures shown in this work, it's hard to believe that "inter consistency may also exist in the common background", especially in Banana, Sofa and Pineapple; Guitar, Hammer, Bowl (in supplementary file), etc. So, whether this operation for background really makes sense is a doubt to me.
3. The group "Tree" is removed because SOD methods behave badly on it, but most other data-driven methods

don't take SISM for reference, if these methods also ignore "Tree"? If not, experiments comparison on this dataset may need a supplementary or something.

4. Correctness: Are the claims and method correct? Is the empirical methodology correct?
Somewhat yes.

5. Clarity: Is the paper well written?
Somewhat yes.

6. Relation to prior work: Is it clearly discussed how this work differs from previous contributions?
Somewhat yes

7. Reproducibility: Are there enough details to reproduce the major results of this work?
No

8. Additional feedback, comments, suggestions for improvement and questions for the authors:
Some of my concerns are addressed. This work need to add some ablation study in manucsript. In addition, the author have not clearly explain the question 3. Therefore, i will keep my rating.

9. Please provide an "overall score" for this submission.
5: Marginally below the acceptance threshold.

10. Please provide a "confidence score" for your assessment of this submission.
5: You are absolutely certain about your assessment. You are very familiar with the related work.

11. Have the authors adequately addressed the broader impact of their work, including potential negative ethical and societal implications of their work?
Only partially, more discussion is needed.

Reviewer #3

Questions

1. Summary and contributions: Briefly summarize the paper and its contributions.
This paper proposes an Intra-saliency Correlation Network (ICNet) to extract intra-saliency cues from the single image saliency maps (SISMs) and obtain inter-saliency cues by correlation techniques. Specifically, the authors first adopt a normalized masked average pooling (NMAP) technique to extract latent intra-saliency categories from the SISMs and semantic features as intra cues. Then, they design a correlation fusion module (CFM) to obtain inter cues by exploiting correlations between the intra cues and single-image features. Besides, the authors also propose a category-independent rearranged self-correlation feature (RSCF) strategy to further improve the Co-SOD performance.

2. Strengths: Describe the strengths of the work. Typical criteria include: soundness of the claims (theoretical grounding, empirical evaluation), significance and novelty of the contribution, and relevance to the NeurIPS community.
+ The strategy of using the correlation to model inter-saliency sounds good and reasonable. It is a natural choice to model inter-saliency patterns across the image group.
+Good experimental results are obtained.

3. Weaknesses: Explain the limitations of this work along the same axes as above.
-It seems that the proposed learning system would be influenced by the number of images in each image group. Th authors are suggested to discuss how big the influence would be.
-As the SISMs of this method are obtained from the most recent SOD methods, it is not so fair to compare this approach to the existing methods.
-It is not clear why Rearranged SCF works.
-Although the whole idea is sound, the proposed approach combines many existing techniques, such as off-the-shelf SOD method, NMAP, non-local feature correlation, etc. This hurts the technical novelty of this work.

4. Correctness: Are the claims and method correct? Is the empirical methodology correct?
Most claims look correct and the methodology is reasonable.

5. Clarity: Is the paper well written?

The paper is well written and easy to follow.

6. Relation to prior work: Is it clearly discussed how this work differs from previous contributions?
The difference between this work and the previous works is clearly discussed.

7. Reproducibility: Are there enough details to reproduce the major results of this work?
No

8. Additional feedback, comments, suggestions for improvement and questions for the authors:
----------------------------update after rebuttal--------------------------
Very appreciate for the authors' efforts to provide the response. Some of my concerns are addressed while some are still remind. Thus, I keep my original rating.

9. Please provide an "overall score" for this submission.
6: Marginally above the acceptance threshold.

10. Please provide a "confidence score" for your assessment of this submission.
5: You are absolutely certain about your assessment. You are very familiar with the related work.

11. Have the authors adequately addressed the broader impact of their work, including potential negative ethical and societal implications of their work?
Only partially, more discussion is needed.

Reviewer #4

Questions

1. Summary and contributions: Briefly summarize the paper and its contributions.
This paper develops a co-saliency detection method that exploits intra- and inter-saliency cues. It integrates intra-saliency features of single image saliency maps (SISMs) and designs a correlation fusion module (CFM) to exploit their correlations. A rearranged self-correlation feature (RSCF) strategy is proposed to obtain robust co-saliency features from inter-saliency cues. The experiments and ablation studies on three co-saliency benchmarks demonstrate effectiveness of intra- and inter-saliency cues, as well as the proposed modules.

2. Strengths: Describe the strengths of the work. Typical criteria include: soundness of the claims (theoretical grounding, empirical evaluation), significance and novelty of the contribution, and relevance to the NeurIPS community.
This paper is well motivated from the observation that salient object detection methods achieve comparable performances over co-saliency methods. Utilizing SISMs for better intra-saliency cue extraction is a novel contribution.

The RSCF strategy can effectively improve the consistency between category-independent co-salient attention maps and category-related image features.

Experimental results are adequate and comprehensive.

3. Weaknesses: Explain the limitations of this work along the same axes as above.
This paper adopts normalized masked average pooling (NMAP) from [23] and dense correlations from [28]. Thus, the main technical novelties is the SCF and its rearranged variant.

While the author claims that directly integrating SISMs generated by any off-the-shelf SOD model is better than taking SISMs as the training targets, it is still unclear why the former is better since both of them utilize inaccurate SISMs. It is also unclear why adding pooling/normalization layers for the latter cannot dilute the inaccuracy.

Though the improvement of MAE is significant, the F-measure and S-measure on MSRC and iCoseg is marginal for ICNet. It is suggested to include an analysis about the possible reasons that degrade the model performance on these metrics.

4. Correctness: Are the claims and method correct? Is the empirical methodology correct?

Both the claim and the method is correct.

5. Clarity: Is the paper well written?

The writing of this paper is good.

6. Relation to prior work: Is it clearly discussed how this work differs from previous contributions?

This work clearly discussed how it is different from previous works.

7. Reproducibility: Are there enough details to reproduce the major results of this work?

Yes

9. Please provide an "overall score" for this submission.

6: Marginally above the acceptance threshold.

10. Please provide a "confidence score" for your assessment of this submission.

3: You are fairly confident in your assessment. It is possible that you did not understand some parts of the submission or that you are unfamiliar with some pieces of related work. Math/other details were not carefully checked.

11. Have the authors adequately addressed the broader impact of their work, including potential negative ethical and societal implications of their work?

Yes