
CoADNet: Collaborative Aggregation-and-Distribution Networks for Co-Salient Object Detection (*Supplementary Material*)

Anonymous Author(s)

Affiliation

Address

email

1 Details of Benchmark Datasets

2 We conduct extensive experiments on four prevailing CoSOD benchmark datasets, including MSRC
3 [4], iCoseg [1], Cosal2015 [6], and CoSOD3k [2]. MSRC [4] is the first available dataset for co-
4 saliency detection, which includes 7 groups of 210 images after removing the non-salient *grass* group,
5 and each group contains 30 images. iCoseg [1] is originally designed for image co-segmentation but
6 is also widely used for co-saliency detection, which consists of 643 images distributed in 38 groups.
7 The group size varies from 4 to 42. In this dataset, the co-salient objects within one group usually
8 share similar background context and object appearances with variation of sizes and poses, and thus
9 is viewed as a relatively simple dataset. Cosal2015 [6] is the first large-scale dataset for co-saliency
10 detection and contains 2,015 images from 50 groups, and each group including 26~52 images. This
11 dataset is much more difficult, and is characterized by complex contextual information, occlusion,
12 large appearance variation, and background clutters. CoSOD3k [2] is the latest and largest dataset
13 carefully tailored for the task of co-salient object detection, which consists of 3,316 images divided
14 in 160 groups. Compared with the other three CoSOD benchmark datasets, CoSOD3k makes a
15 significant leap in terms of sufficient group semantic categories, diverse object sizes and appearances.
16 Hence, it is considered as the most challenging CoSOD benchmark dataset.

17 2 Experiments

18 In this section, we provide more visual examples (*i.e.*, the banana, popsicle, pepper, and bowl groups)
19 in Fig. 1, the P-R curves of different competing methods on four datasets in Fig. 2, and the detailed
20 flowchart of baseline model is shown in Fig. 3.

21 From Fig. 1, we can observe that the proposed CoADNet produces more visually competitive results
22 with accurate localization, complete structure, and high consistency of co-salient objects. Moreover,
23 our method can effectively suppress the cluttered background regions. As shown in Fig. 2, we can see
24 that our method with the Dilated ResNet-50 [5] backbone (*i.e.*, the red solid line) achieves the highest
25 precision on all datasets except the smallest MSRC dataset. Overall, both qualitative and quantitative
26 results in the main body and supplementary material demonstrate the superiority and effectiveness of
27 our method.

28 For clarity, in this *Supplementary Material*, we also intuitively provide the structure of the baseline
29 model in Fig. 3. Note that, for fair comparison, the backbone feature extractor in baseline model is
30 pretrained on a single-image saliency detection dataset (*i.e.*, DUTS [3]) for shared feature learning.



Figure 1: Visual examples of co-saliency maps generated by different methods. Our method with the Dilated ResNet-50 backbone is marked with the yellow shading.

31 References

- 32 [1] D. Batra, A. Kowdle, D. Parikh, J. Luo, and T. Chen. iCoseg: Interactive co-segmentation with
33 intelligent scribble guidance. In *Proc. CVPR*, pages 3169–3176, 2010.
- 34 [2] D.-P. Fan, Z. Lin, G.-P. Ji, D. Zhang, H. Fu, and M.-M. Cheng. Taking a deeper look at the
35 co-salient object detection. In *Proc. CVPR*, 2020.
- 36 [3] L. Wang, H. Lu, Y. Wang, M. Feng, D. Wang, B. Yin, and X. Ruan. Learning to detect salient
37 objects with image-level supervision. In *Proc. CVPR*, pages 136–145, 2017.
- 38 [4] J. Winn, A. Criminisi, and T. Minka. Object categorization by learned universal visual dictionary.
39 In *Proc. ICCV*, pages 1800–1807, 2005.

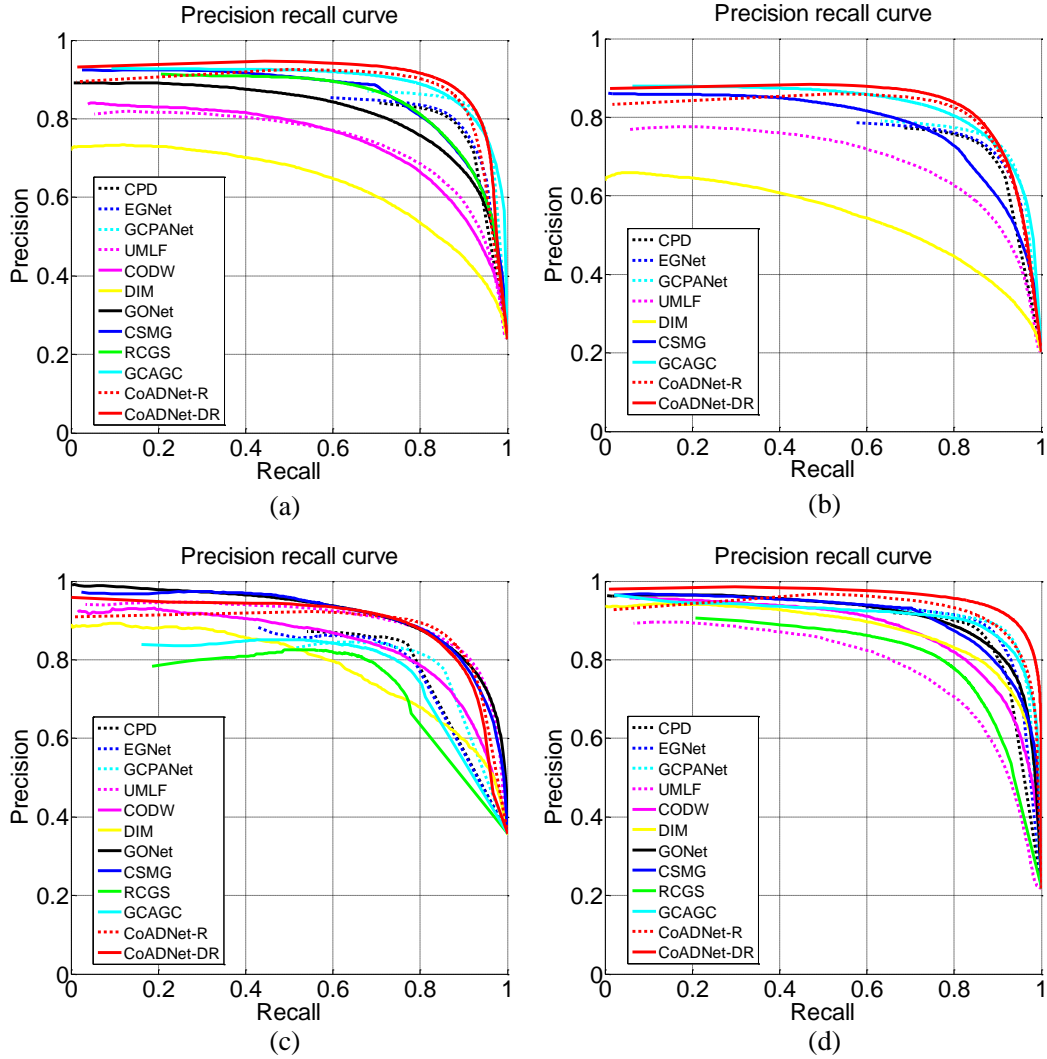


Figure 2: P-R curves of different methods on four datasets. (a)-(d) correspond to the P-R curves on the Cosal2015 dataset, CoSOD3k dataset, MSRC dataset, and iCoseg dataset.

- 40 [5] F. Yu, V. Koltun, and T. Funkhouser. Dilated residual networks. In *Proc. CVPR*, pages 472–480,
 41 2017.
- 42 [6] D. Zhang, J. Han, C. Li, and J. Wang. Co-saliency detection via looking deep and wide. In *Proc.*
 43 *CVPR*, pages 2994–3002, 2015.

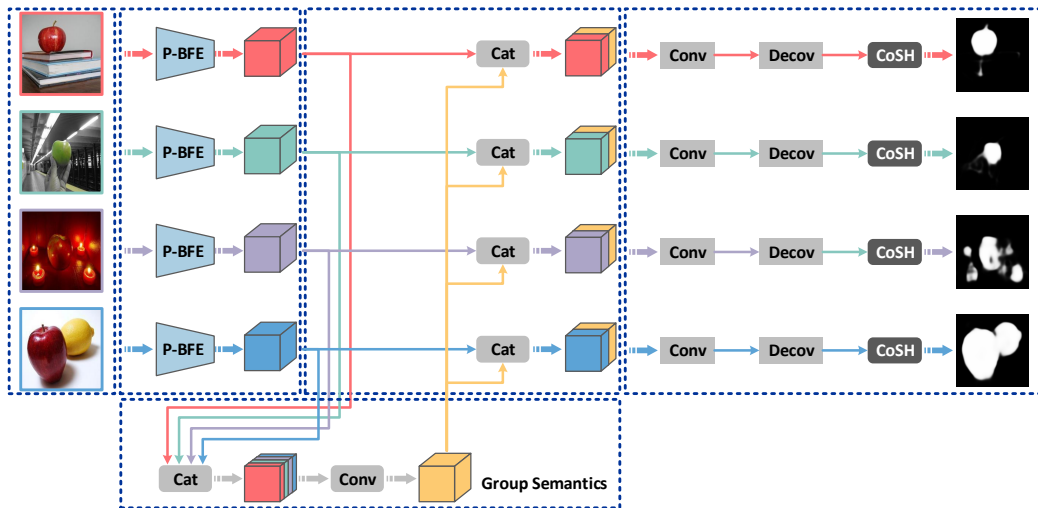


Figure 3: The overall framework of our baseline model constructed for ablative analysis. The input group of images pass through the backbone feature extractor pretrained (P-BFE) on a single-image saliency detection dataset for shared feature learning. The generated individual image features are sequentially concatenated and convolved into group semantic representations, which are further equally concatenated back with different individuals to obtain co-saliency features. During feature decoding, we apply three cascaded deconvolutional layers to produce full-resolution co-saliency features, which are fed into a co-saliency head (CoSH) to predict co-saliency maps.