A. Appendix

In this supplementary material, we analyze our method with some visualization results and provide more qualitative comparisons between our model and other state-of-the-art methods. Then, we provide more experimental results in simulated real scenarios. Additionally, we analyze the difference between our GEM with different settings of k and randomly exchanged images with quantity settings of e.

A.1. Group Exchange-masking Analysis

We display some response maps in the decoding phase with or without (w/o) the designed GEM in Figure 1. We can see that with GEM, our model can recognize the noisy images well, while w/o GEM the model incorrectly segments the non-co-salient objects. Because of the training process with noisy images, our model also has good discrimination ability to distinguish the interfering objects in a single image, which is better than that w/o GEM.

We further analyze the influence of k. The number of noisy images is chosen to be less than the number of remaining relevant images in the group, so that the co-salient object in the noisy images forms a negative object but not the dominant co-salient object. Our group size is set to 5, and we set k to 1 and 2 for experiments on CoCA [5]. The results are shown in Table 1, and when we set k to 2, MAE, E_{ϕ}^{max} and F_{β}^{max} become worse. This means that in our scheme, moderate noisy images addition can improve model performance. When there is too much noise, it will bring a great challenge to the model to extract common information, further affecting the model performance. Based on this, we think the best proportion of noisy images in one group is about 20%, that is, when the group size becomes larger, the number of noise images should also be increased to ensure good performance. In addition, we also conduct experiments on the random exchange number e as shown in Table 2. We can see that the effect of randomly exchanging is not as good as top-k exchanging in general.

A.2. More Qualitative Comparisons

On the basis of the existing largest test set [1], we build several sub data sets to simulate the real scenes. Specifically, in each existing group, we mix several pictures of different categories, the number of which is smaller than the images in the group. We set two scenarios in which the number of co-salient objects and the number of non-co-salient interference objects are 4:1 and 3:2 as shown in Figure 2 and Figure 3, corresponding to the difficulty from low to high. Additionally, we also conduct the experiments on interference objects from different categories as shown in Figure 4. From Figure 2, we can see that when there are low proportion of noisy pictures in the picture group, our method can distinguish the noisy images well. However, other methods yield more or less wrong segmentation results. As shown in Figure 3, when the proportion of the noisy image further increases, the regions wrongly segmented by other methods also become larger, but our method can still accurately distinguish the noisy images. We also consider the case that noisy images come from different categories. Figure 4 shows the advantages of our method over other methods that cannot accurately identify the non-co-salient objects.

Table 1. Comparisons of different k settings on CoCA.

Setting of k	$MAE \downarrow$	$S_{\alpha} \uparrow$	$E_{\phi}^{max} \uparrow$	$F_{\beta}^{max} \uparrow$
0	0.104	0.724	0.802	0.597
1	**0.095**	0.726	**0.808**	**0.599**
2	0.100	**0.732**	0.793	0.578

Table 2. Comparisons of different e settings on CoCA.

Setting of e	$MAE \downarrow$	$S_{\alpha} \uparrow$	$E_{\phi}^{max} \uparrow$	$F_{\beta}^{max} \uparrow$
1	0.101	0.712	0.809	0.582
2	0.105	0.710	0.788	0.572

References

[1] Deng-Ping Fan, Zheng Lin, Ge-Peng Ji, Dingwen Zhang, Huazhu Fu, and Ming-Ming Cheng. Taking a deeper look at co-salient object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2919–2929, 2020. 1

[2] Qi Fan, Deng-Ping Fan, Huazhu Fu, Chi-Keung Tang, Ling Shao, and Yu-Wing Tai. Group collaborative learning for co-salient object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12288–12298, 2021. 2, 3

[3] Siyue Yu, Jimin Xiao, Bingfeng Zhang, and Eng Gee Lim. Democracy does matter: Comprehensive feature mining for co-salient object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 979–988, 2022. 2, 3

[4] Ni Zhang, Junwei Han, Nian Liu, and Ling Shao. Summarize and search: Learning consensus-aware dynamic convolution for co-saliency detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4167–4176, 2021. 2, 3

[5] Zhao Zhang, Wenda Jin, Jun Xu, and Ming-Ming Cheng. Gradient-induced co-saliency detection. In *Proceedings of the European Conference on Computer Vision*, pages 455–472. Springer, 2020. 1, 2, 3

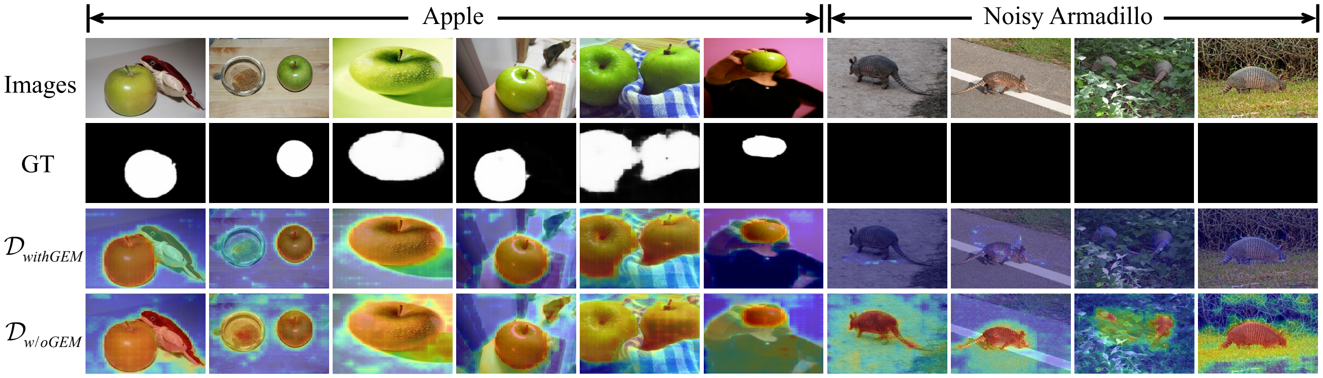

Figure 1. The response maps with or w/o GEM. The first row shows image group with noisy samples that come from another category. The second row is the ground truth. The third and forth rows show the response maps $\mathcal{D}_{withGEM}$ and $\mathcal{D}_{w/oGEM}$ from the decoder.

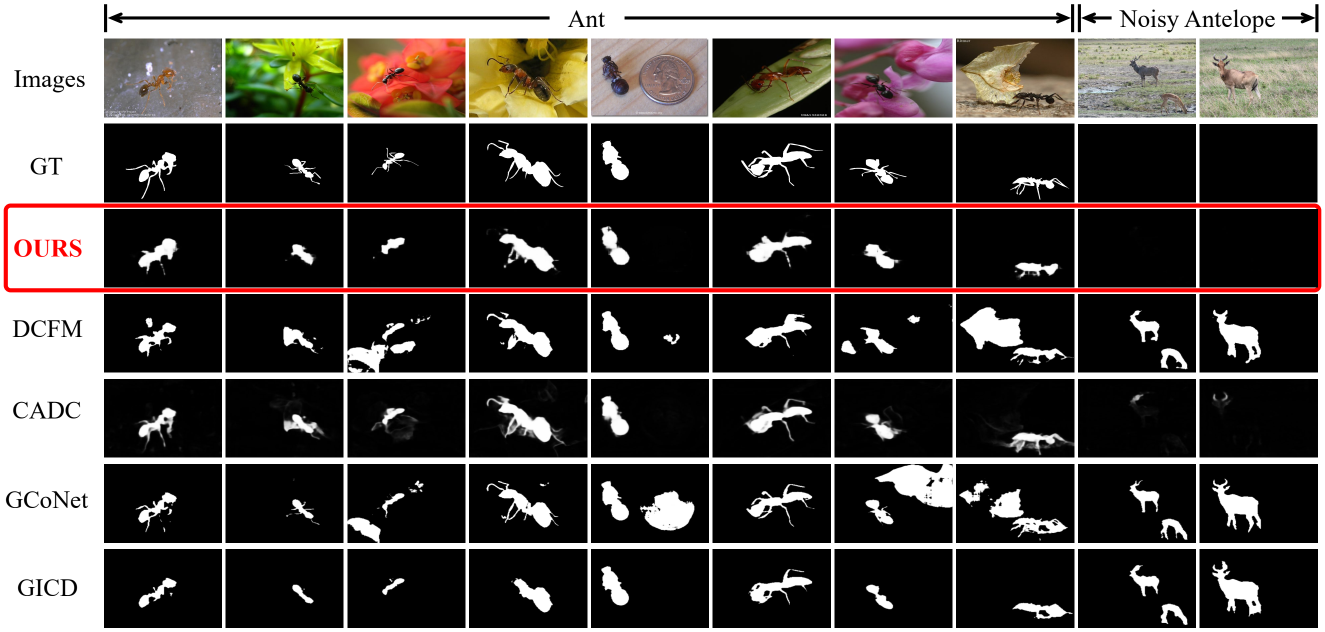

Figure 2. More visualizations of our method and the compared recent state-of-the-art methods, including DCFM [3], CADC [4], GCoNet [2] and GICD [5].

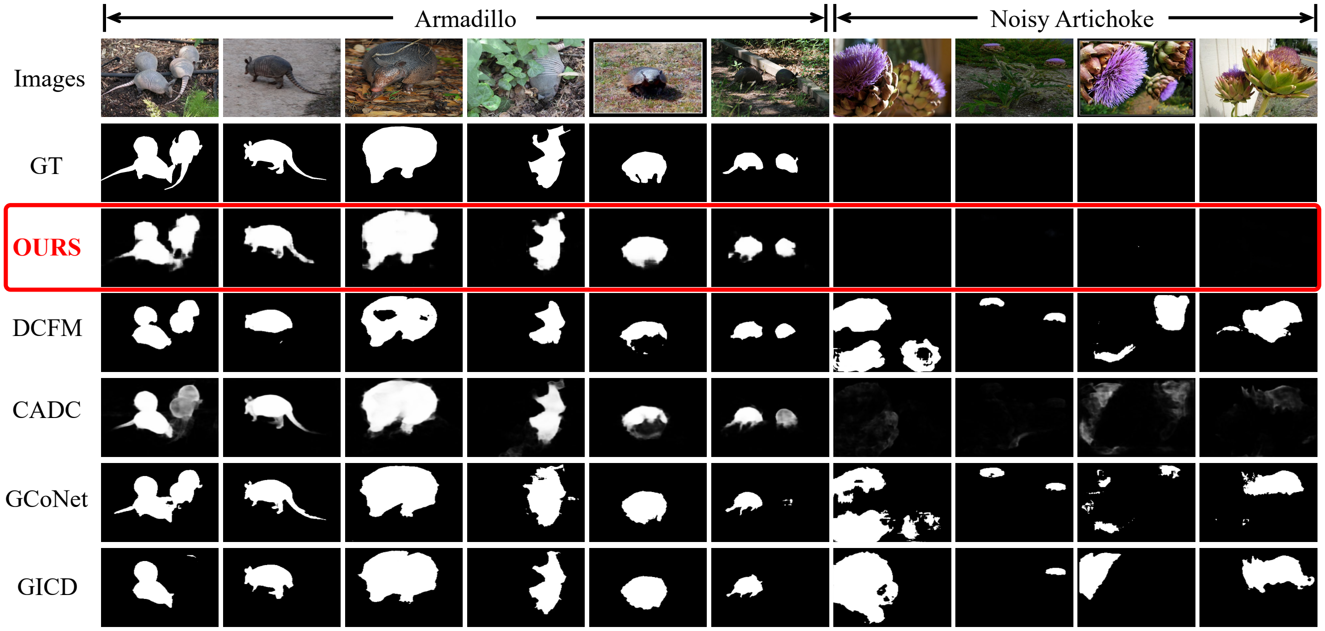

Figure 3. More visualizations of our method and the compared recent state-of-the-art methods, including DCFM [3], CADC [4], GCoNet [2] and GICD [5].

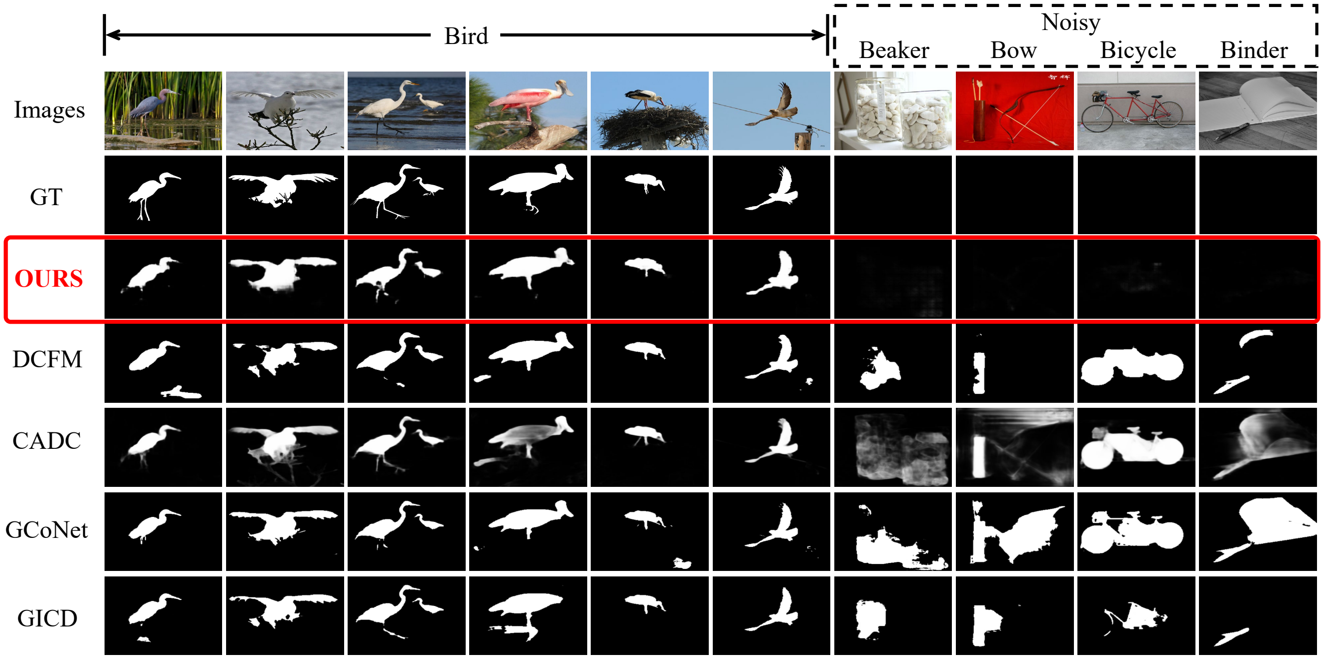

Figure 4. More visualizations of our method and the compared recent state-of-the-art methods, including DCFM [3], CADC [4], GCoNet [2] and GICD [5].