Simultaneously Localize, Segment and Rank the Camouflaged Objects
–*Supplementary Material*–

Yunqiu Lv[1,‡] Jing Zhang[2,3,‡] Yuchao Dai[1]✉ Aixuan Li[1] Bowen Liu[1] Nick Barnes[2] Deng-Ping Fan[4]
[1] Northwestern Polytechnical University, China [2] Australian National University, Australia
[3] CSIRO, Australia [4] Inception Institute of AI (IIAI), Abu Dhabi, UAE
‡ Equal contributions; ✉ Corresponding author:daiyuchao@nwpu.edu.cn

Abstract

In this supplementary material, we provide more details about our dataset and experiments, which include: 1) More details about the dataset collection and detailed visualization of our camouflage ranking based and discriminative region based dataset in Section 1; 2) Analysis of the proposed testing dataset in Section 2; 3) More visual results of our discriminative region localization model, camouflaged object detection model and ranking model in Section 3 and 4) Additional details about the implementation of our ranking based experiments in Section 4.

1. Dataset Collection and Visualization

Our dataset is generated by employing an eye tracker, a common device in psychological research of attention, visual perception, and reading. We invite six observers in total to participate in the eye-tracking experiment and all observers have never seen these images before the experiment. The dataset is shuffled randomly and partitioned into 23 groups and each has 100 images (except a group with 80 images). Each image is resized to the same size as the screen resolution, 1920×1080 pixels. The observers stop to have a rest after every 25 minutes of performing the task. Each recording group begins with a five-point calibration. Before each target image, a black screen is shown for 2 seconds to avoid influence by the previous image. During collection, observers were required to find all the camouflaged objects in an image and switch to the next manually once they believe all instances have been observed. The maximal delay for people to view an image is set to 20s in case they fail to find any objects at all.

The recorded eye fixations indicate the attention shift of the observers. Accordingly, we collect the fixation maps of all observers to build the discriminative region based dataset, where the positions with strong responses within an instance indicate the discriminative regions, and use the time of each fixation to compute the detection delay, which is the main indicator for our ranking based dataset. For better visualization, Fig. 1 provides the original images, colorised ranking maps and fixation maps for camouflage rank 3 (easiest, green), rank 2 (median, orange) and rank 1 (hardest, blue), respectively.

In this way, with the original instance level annotation, our relabeled dataset has five types of annotation as shown in Fig. 2, including binary camouflaged object annotations, edge annotations, instance annotations, discriminative region annotations, gray-scale ranking annotations.

2. Testing Dataset Analysis

In this paper, we contribute a large-scale testing dataset, NC4K, for camouflaged object detection.

As the scale and location of the camouflaged objects are key factors that influence the accuracy of detection, we summarize scale and location distribution of the camouflaged objects in the new testing set and existing testing set as shown in Fig. 3. We also show the furthest camouflaged point to image center, which serves as an effective representation of the location of the camouflaged objects [2]. The scale distribution of the camouflaged objects show that we have a wide location distribution of camouflaged objects, and relatively smaller scales of camouflaged objects, which make our testing set a challenging dataset to evaluate model generalization ability.

Since some complex camouflage strategies pose a great challenge to the annotators, we filtered out wrongly-annotated data from our dataset and relabeled them for 3 times. Therefore, high-quality annotations are provided in the NC4K testing dataset. In Fig. 4, we list 6 challenging cases in camouflage annotation: 1) The object in the background is included as a part of the camouflaged object (Redundant Object); 2) The camouflaged object is labeled incompletely (Incomplete Object); 3) Multiple objects are fused as a single object (Fused Object); 4) The camouflaged object is ignored by the annotator (Missing

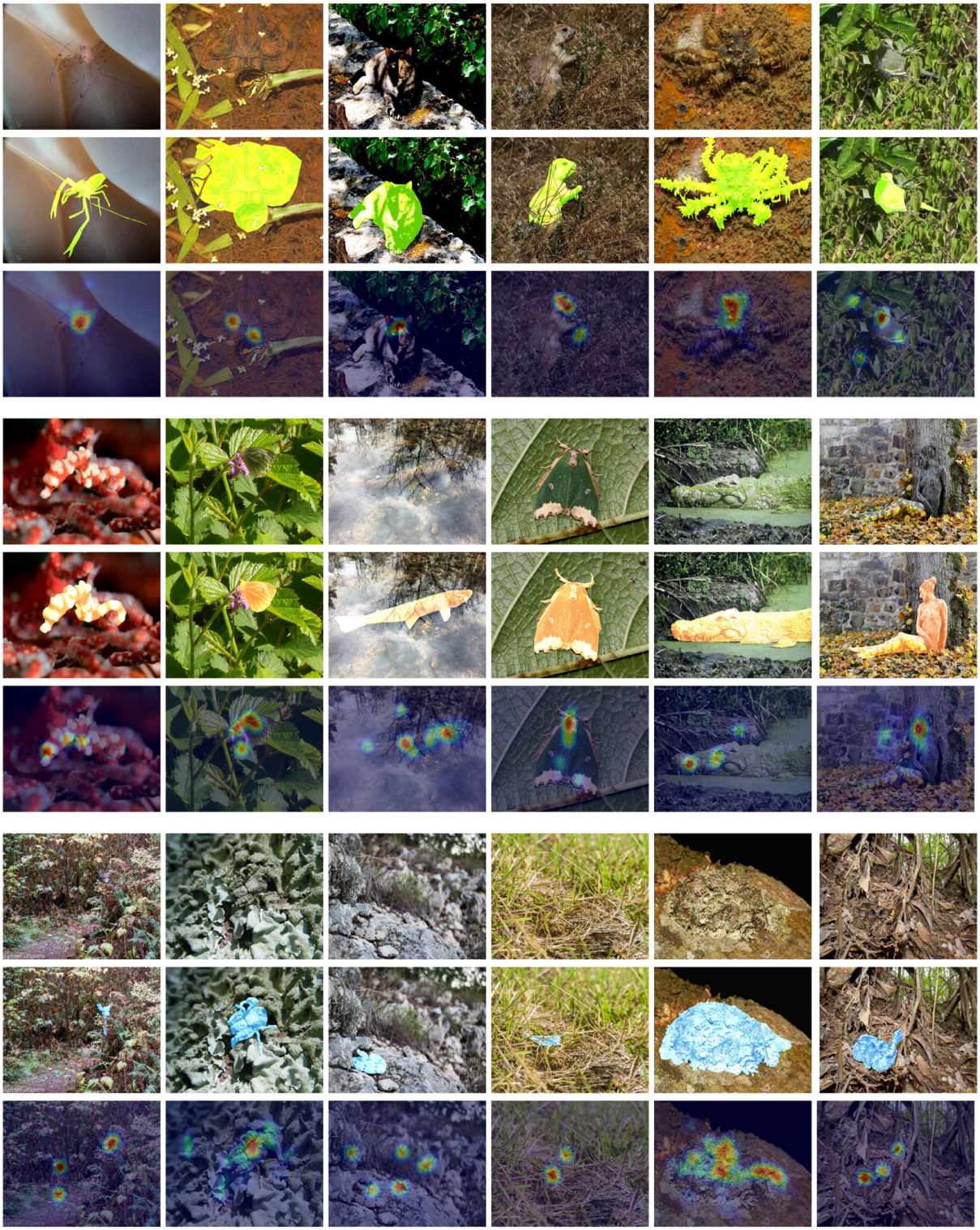

Figure 1: Visualization of the ranking based and discriminative region based dataset. For each rank, the original image, ranking maps and fixation maps are listed. Green, orange and blue represent camouflage rank 3 (easiest), rank 2 (median) and rank 1 (hardest).

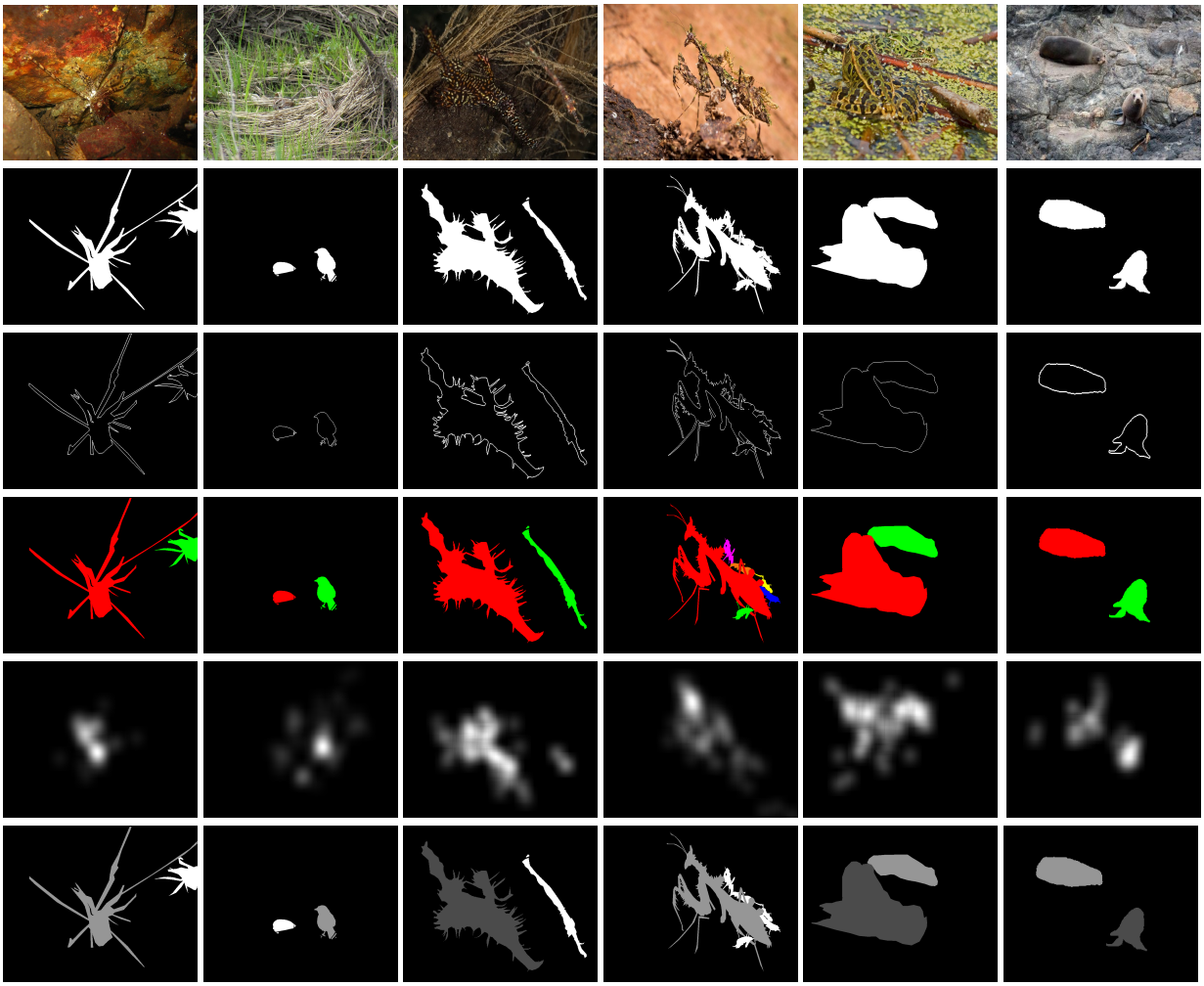

Figure 2: Visualization of annotations that our current relabeled camouflage dataset provided. From top to bottom: the original images, binary camouflaged object annotations, edge annotations, instance annotations, discriminative region annotations, gray-scale ranking annotations.

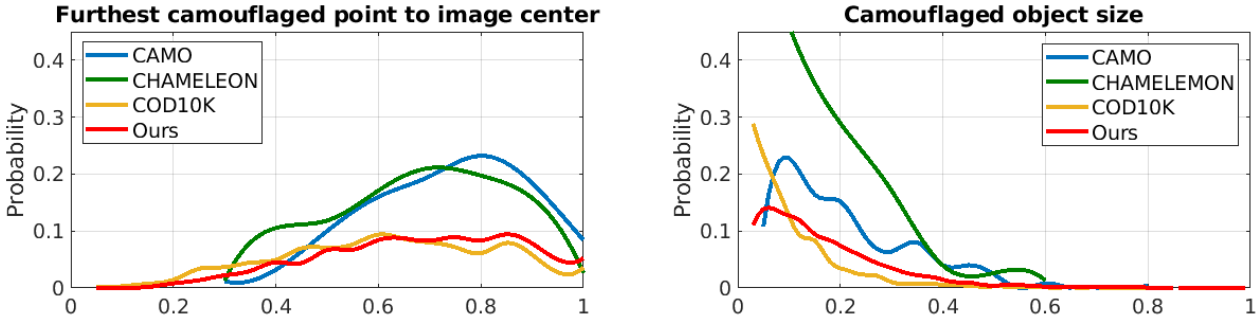

Figure 3: Analysis of the new testing dataset

Object); 5) A single object is separated into several instances (Separated Annotation); 6) The non-camouflaged object is labeled while the camouflaged one is missed (Misidentification). These problems are carefully avoided in our annotations.

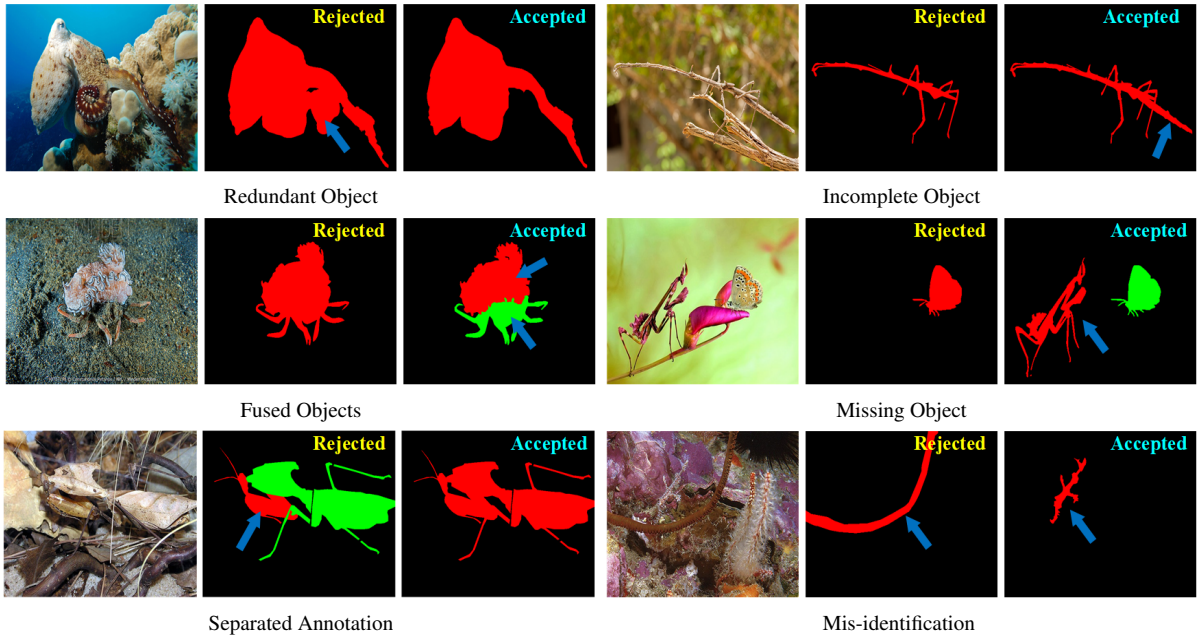

Figure 4: Visualization of six challenges in annotations of our NC4K testing dataset.

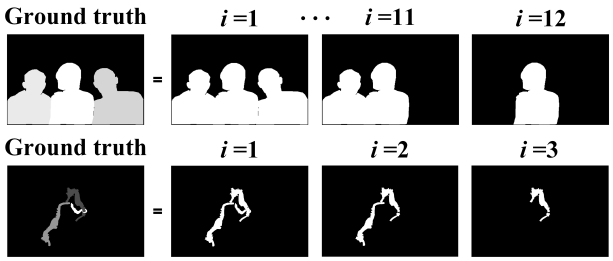

Figure 5: Stacked maps of the saliency ranking ground truth (the top row) and camouflage ranking ground truth (the bottom row) used in RSDNet [1]. The lighter color in the original ground truths represents the higher level of saliency and camouflage, respectively.

3. Prediction Visualization

In order to illustrate the effectiveness of our proposed method, more visual results are shown in Fig. 6 by comparing the prediction with the ground truth. As we can observe from Fig. 6, the proposed model could segment the camouflaged objects, predict the discriminative region and rank the level of camouflage simultaneously.

4. Implementation Details

Since we propose the first camouflage ranking model, we compare it with other ranking models for different tasks, including RSDNet [1], MS-RCNN [3] and SOLOv2 [4]. RSDNet is a saliency ranking model, which evaluates the rank of each instance according to their saliency. We rearrange the experimental setting by following the assumption that the higher level of saliency corresponds to the lower level of camouflage. In order to learn the progressive relationship of the ranks, RSDNet splits the ground truth saliency map into a set of stacked binary maps, where the i-th map contains the instances that at least i observers labeled them as salient ones. Therefore, the map contains all instances when $i = 1$ and only contains the most salient instances when $i = N$, where N denotes the number of ranks. Fig. 5 shows the stacked maps of saliency ranking and camouflage ranking, respectively. In the camouflaged ranking ground truth, the brighter value denotes the higher camouflage rank, while in the saliency ranking ground truth, lighter color denotes the higher level of saliency. Since most images in the saliency ranking dataset contain multiple instances, RSDNet computes saliency of each instance, and then it produces the instance-level ranking based on the mean saliency of each instance, *e.g.* the higher the mean saliency, the higher rank of saliency. However, images in our camouflage ranking dataset usually have only one instance. Instead of computing the mean saliency of each instance, we set 3 ranges for it, *e.g.* [151,255], [86,150], [25,85], to determine camouflage rank 3 (easiest), rank2 (median) and rank 1 (hardest) based on the contradict attribute of camouflage and saliency. If the saliency response of an instance is less than 25, we consider that RSDNet fails to detect it.

As the model needs to detect each instance for ranking, we compare our ranking model with MS-RCNN, which is also based on Mask-RCNN, and SOLOv2 for instance segmentation. The setting of learning rate (5e-5), batch-size

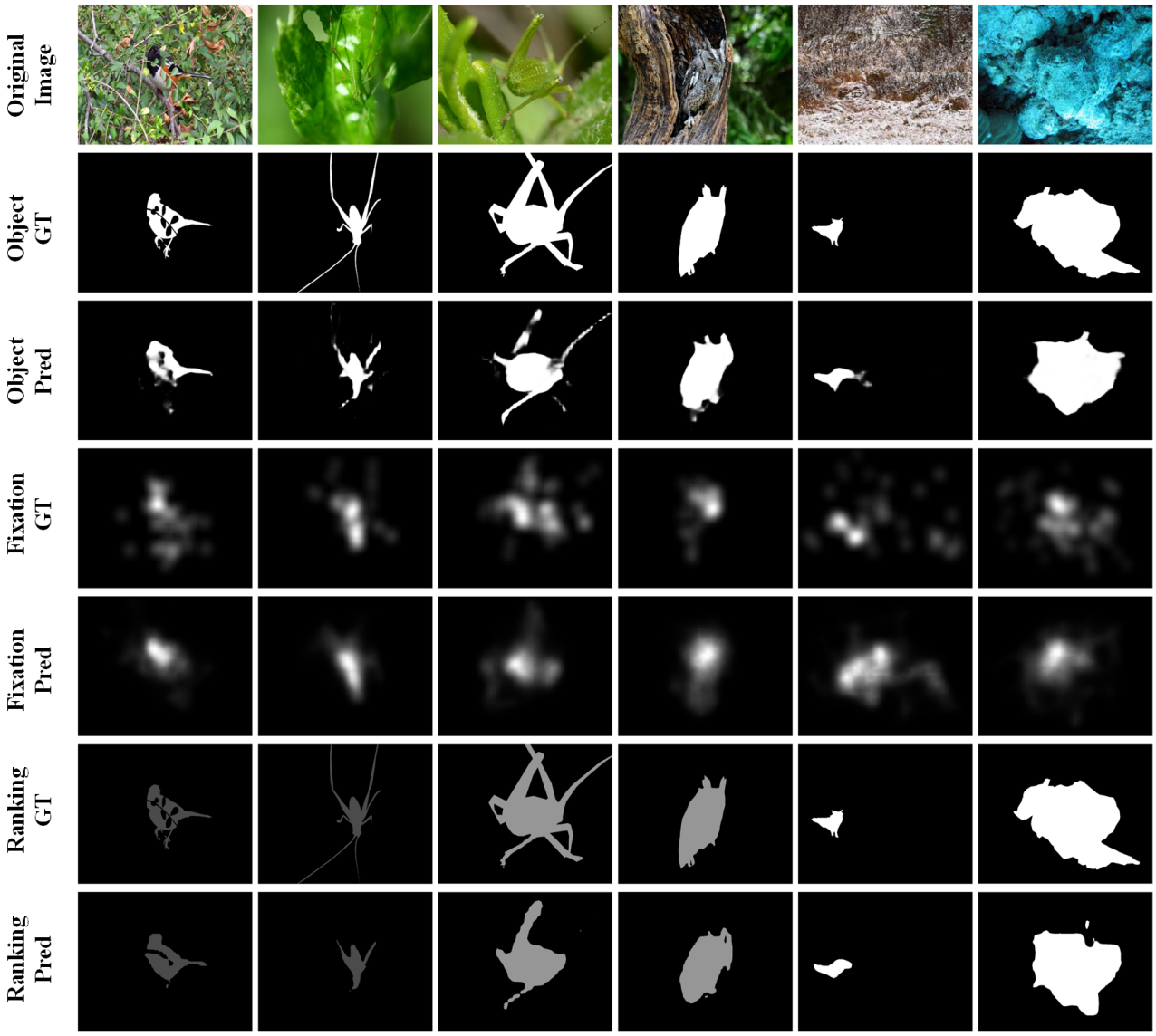

Figure 6: Visual results predicted by the proposed model. From top to bottom: the original image, ground truth of camouflaged object detection, prediction of camouflaged object detection, ground truth of discriminative region prediction, prediction of discriminative region prediction, ground truth of camouflage ranking, prediction of camouflage ranking. In the last two rows, the larger gray value denotes higher level of camouflage.

(10) and iteration (10k) are the same as those in the proposed model.

We visualize the comparison between the proposed ranking model and other ranking based methods in Fig. 7. The proposed method, MS-RCNN [3] and SOLOv2 [4] achieve the segmentation and ranking at the same time, while RS-DNet [1] has to borrow the instance-level ground truth to infer the ranks according to the saliency maps. We observe that the proposed ranking method is able to detect the camouflaged object and give the ranks more precisely than other competing methods, which is consistent with the conclusion in the main paper.

References

[1] Md Amirul Islam, Mahmoud Kalash, and Neil DB Bruce. Revisiting salient object detection: Simultaneous detection, ranking, and subitizing of multiple salient objects. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 7142–7150, 2018. 4, 5, 6

[2] Deng-Ping Fan, Ge-Peng Ji, Guolei Sun, Ming-Ming Cheng, Jianbing Shen, and Ling Shao. Camouflaged object detection. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 2777–2787, 2020. 1

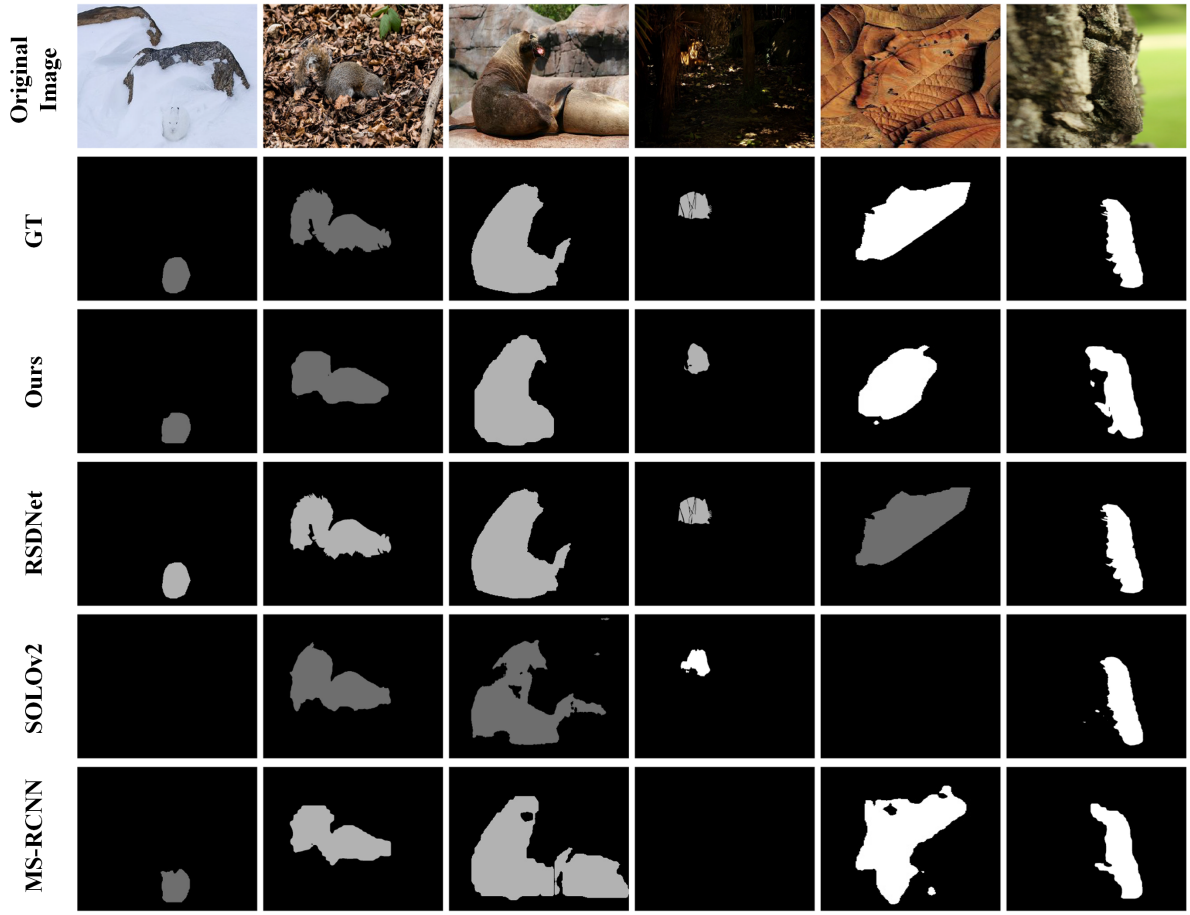

Figure 7: Qualitative comparisons between the proposed model and the competing methods for ranking, including RSDNet[1], SOLOv2[4], MS-RCNN[3].

[3] Zhaojin Huang, Lichao Huang, Yongchao Gong, Chang Huang, and Xinggang Wang. Mask scoring r-cnn. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6409–6418, 2019. 4, 5, 6

[4] Xinlong Wang, Rufeng Zhang, Tao Kong, Lei Li, and Chunhua Shen. Solov2: Dynamic, faster and stronger. *arXiv preprint arXiv:2003.10152*, 2020. 4, 5, 6