

# Supplementary Materials: The Name of the Title is Hope

Anonymous Authors

## 1 APPENDIX

We first provide more training and testing details (Appendix A). We then experiment with an important hyperparameter  $\lambda$  in adaptive query selection. And we report the performance of AQSFormer with multi-scale instances on the standard benchmark set of NC4K (Appendix B). Finally, we provide more visualization results of the ablation studies (Appendix C).

## 2 APPENDIX A

### 2.1 Training Details

During the training process, we follow the settings of Mask2former to calculate the loss function on the sampled K points instead of on the entire prediction map. Specifically, we sample  $112 \times 112$  important points according to the settings in Mask2Former. This step is not required during testing.

## 3 APPENDIX B

### 3.1 Hyperparameter $\lambda$

Entropy and variance play important roles in activity evaluation of the queries. However, due to numerical differences, the direct summation method is not the best solution. Therefore, we set a hyperparameter  $\lambda$  to balance the relationship between the two. In Table 1, we show the results under different settings, and finally we set  $\lambda$  to 10.

### 3.2 Results of Different Scales

More importantly, we compare the results of the model on objects of different sizes, as shown in the Table 2. We compare the camouflaged instance prediction scores (e.g.,  $AP_m$  and  $AP_l$ ) at medium and large scales on two benchmark datasets. It can be found that our model improves significantly on large instances than on medium-scale instances. This is mainly the way of boundary positional embedding, which can separate instances to a great extent. It is relatively easier to extract large-scale camouflage instances than medium-scale instances, so it has better performance on large-scale instances.

Table 1: Hyperparameter  $\lambda$  in the activity evaluation of the queries.

$\lambda$	AP	$AP_{50}$	$AP_{75}$
1	47.2	73.1	48.5
3	47.3	73.5	48.9
5	47.6	73.9	49.5
10	<b>48.1</b>	<b>74.3</b>	<b>50.4</b>
15	47.8	74.0	50.1
20	47.4	73.7	48.8

Table 2: Results of camouflaging instances at different scales.

	COD10K		NC4K	
	$AP_m$	$AP_l$	$AP_m$	$AP_l$
Mask2Former ResNet-50	19.5	47.4	23.4	49.5
OSFormer ResNet-50	22.9	45.3	22.5	45.3
<b>AQSFormer ResNet-50 [Ours]</b>	22.5	50.8	24.1	53.1
<b>AQSFormer ResNet-101 [Ours]</b>	25.2	52.3	28.1	55.3
<b>AQSFormer Swin-Tiny [Ours]</b>	25.1	56.7	31.1	60.0

## 4 APPENDIX C

### 4.1 visualizations of Adaptive Query Selection

To improve the visualization, for each image we only show the selected seven queries in Figure 1. We can find that the selected queries can stably focus on camouflaged instances in images, and do not produce invalid and false positive queries. After obtaining an effective representation, the model can more easily distinguish different instances due to self-attention and FFN layer to model the relationship between queries.

### 4.2 Visualization of More Scenes

In Figure 2, we show more scene prediction results. The following example mainly show extremely difficult case, the instances is almost perfectly embedded in the environment. We can find two points: (i) Our model does not generate redundant and false positive predictions. This is mainly achieved with the adaptive query selection strategy, which filters out many useless queries. (ii) Our model has better discriminative ability in occluded and overlapping instances. This also verifies that the way of encoding the boundary position can improve the insufficiency of the selection strategy on occluded instances, further improving the correctness of the prediction.

### 4.3 Failed Cases

We also show our failure cases in Figure 3. We find that our model often misses camouflaged instances with small structures. There are two main reasons for this. First, the small target itself is more difficult, which is also the pain point of many detection models. Secondly, the feature scale extracted by our boundary position encoding is low, and the features of the captured camouflage instances are less, which exacerbates the difficulty of detecting small targets. By the way, other models have difficulty not only detecting small object regions but also distinguishing different instances.

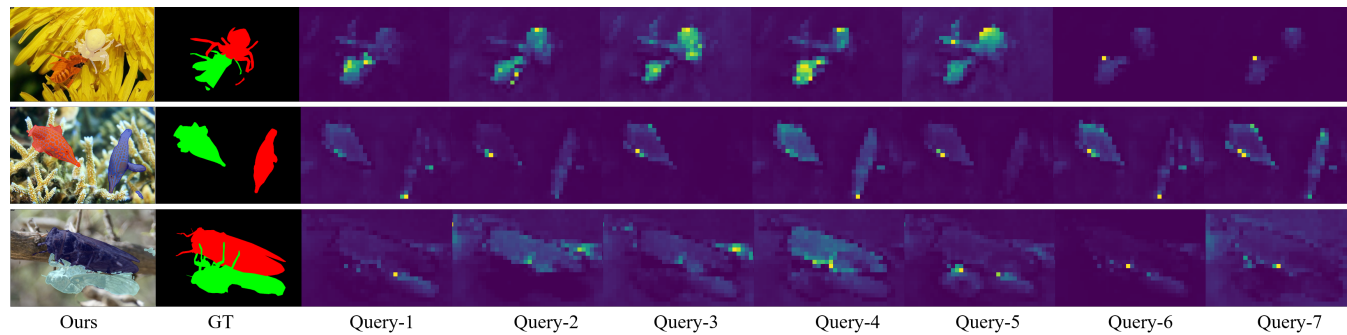


Figure 1: Visualization of the selected queries.

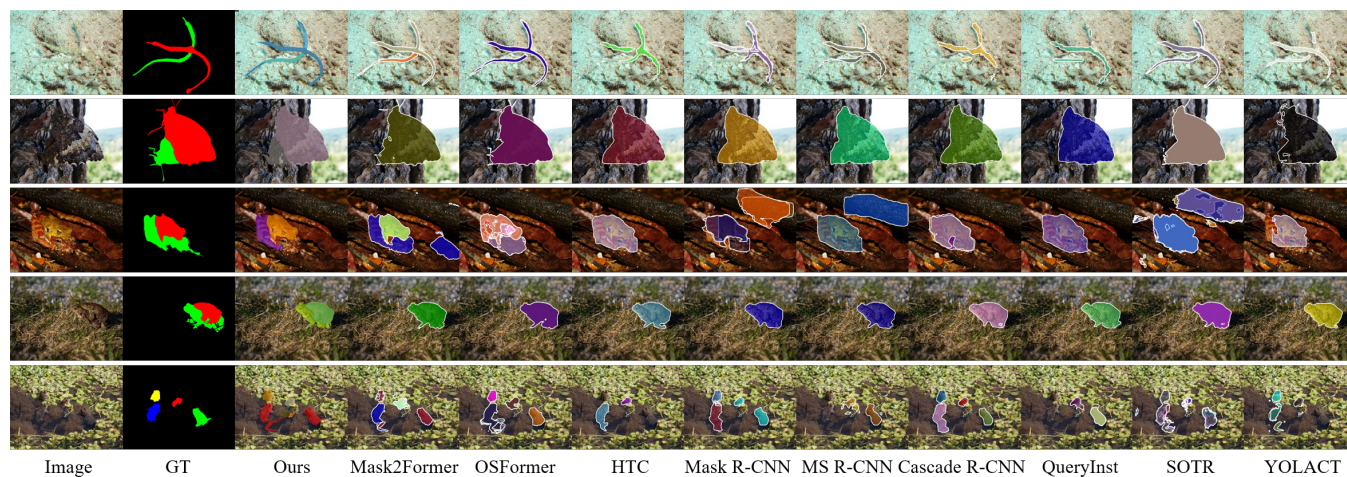


Figure 2: Visual comparison results of our model with other models. We show predictions with confidence scores greater than 0.5.

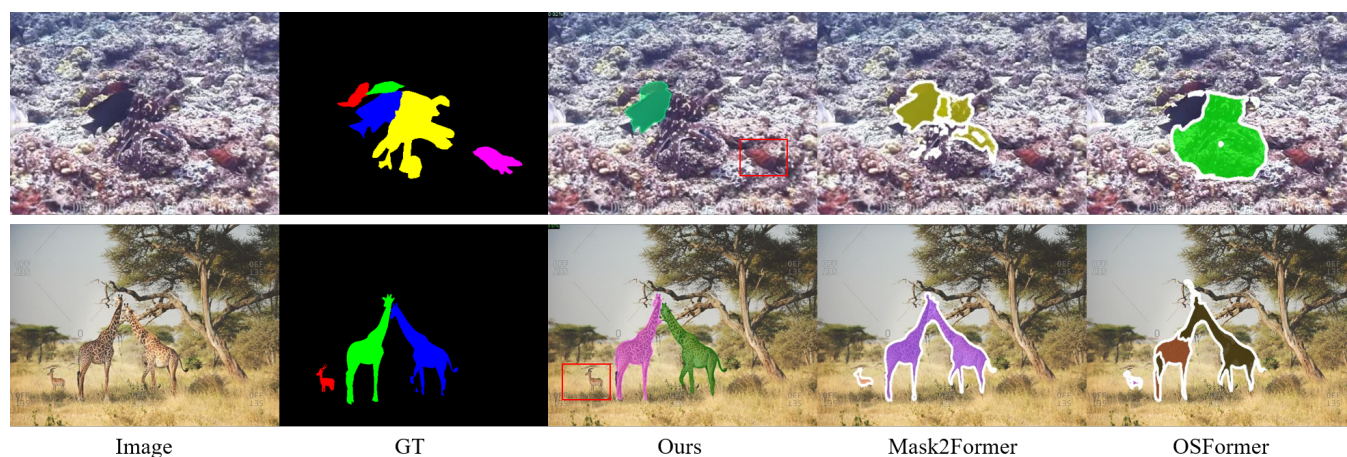


Figure 3: Some failure cases.