Weakly-supervised Object Representation Learning for Few-shot Semantic Segmentation

Supplementary Materials

1. Empirical Study on Challenging Scenarios

We provide empirical analyses of our approach under challenging few-shot segmentation scenarios. To do this, we provide example results on Pascal-5^i dataset under each of the following challenging scenarios:

- **One-to-Many Matching**: The support example has one object and the query image has multiple objects. This scenario requires the object representation extracted from a single object to generalize well such that it can match multiple objects with appearance variations. Example results are visualized in Figure 1.
- **Many-to-One Matching**: The support example has multiple objects and the query image has only one object. This scenario requires the model to be able to aggregate information from multiple objects with varying sizes and appearances. Example results are visualized in Figure 2.
- **Small-to-Large / Large-to-Small Matching**: Objects in the support example are small while objects in the query image are large, or vice versa. This scenario requires the model to be able to effectively produce object representation from object features in different scales. Example results are visualized in Figure 3.
- **Change of Viewing Angles**: When the viewing angle of an object in support image and query image has large variation, it is more difficult to match the objects. Example results are visualized in Figure 4.

2. Failure Cases

We provide example failure cases and the corresponding discussions in Figure 5. The results are generated using our model with ResNet-101 backbone and 513×513 input size.

3. More Example Results on Pascal-5^i Dataset.

More qualitative results of our model on Pascal-5^i dataset are visualized in Figure 6 and Figure 7. All results are generated using our model with ResNet-101 backbone and 513×513 input size.

4. More Example Results on COCO-20^i Dataset.

More qualitative results of our model on COCO-20^i dataset are visualized in Figure 8 and Figure 9. All results are generated using our model with ResNet-101 backbone and 513×513 input size.

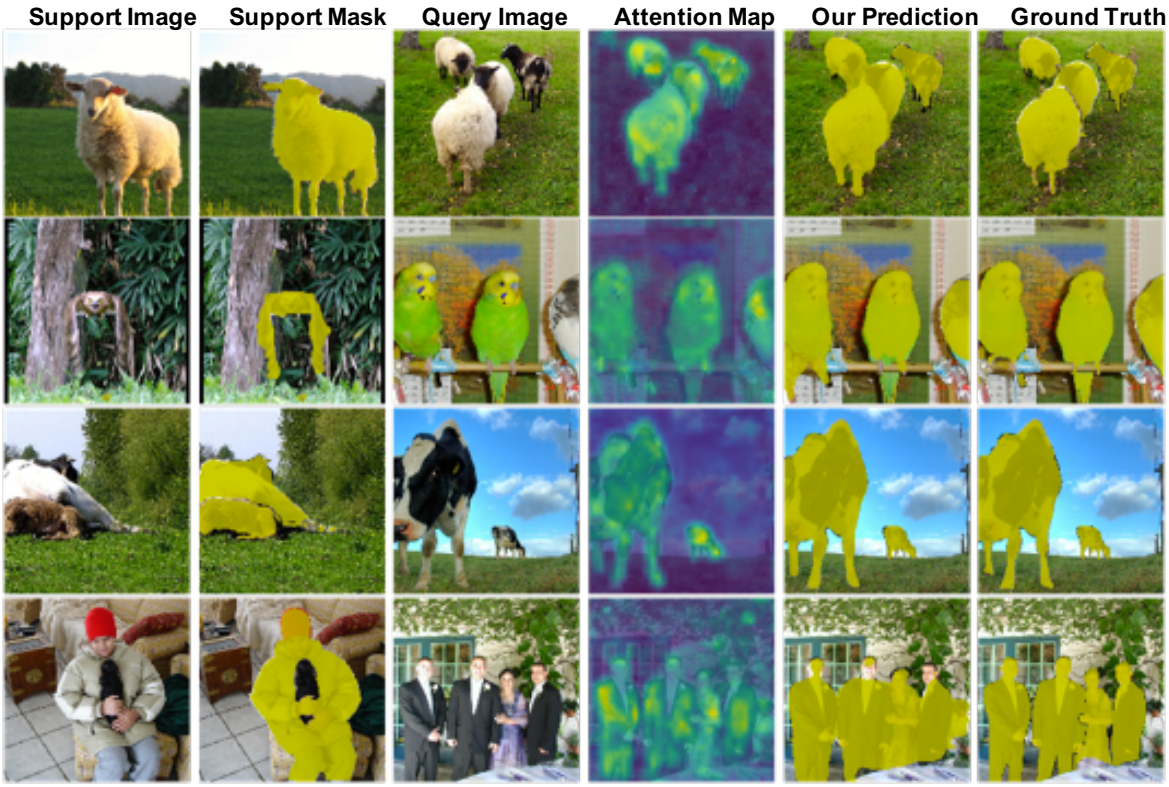

Figure 1. Example results under “one-to-many matching” scenario.

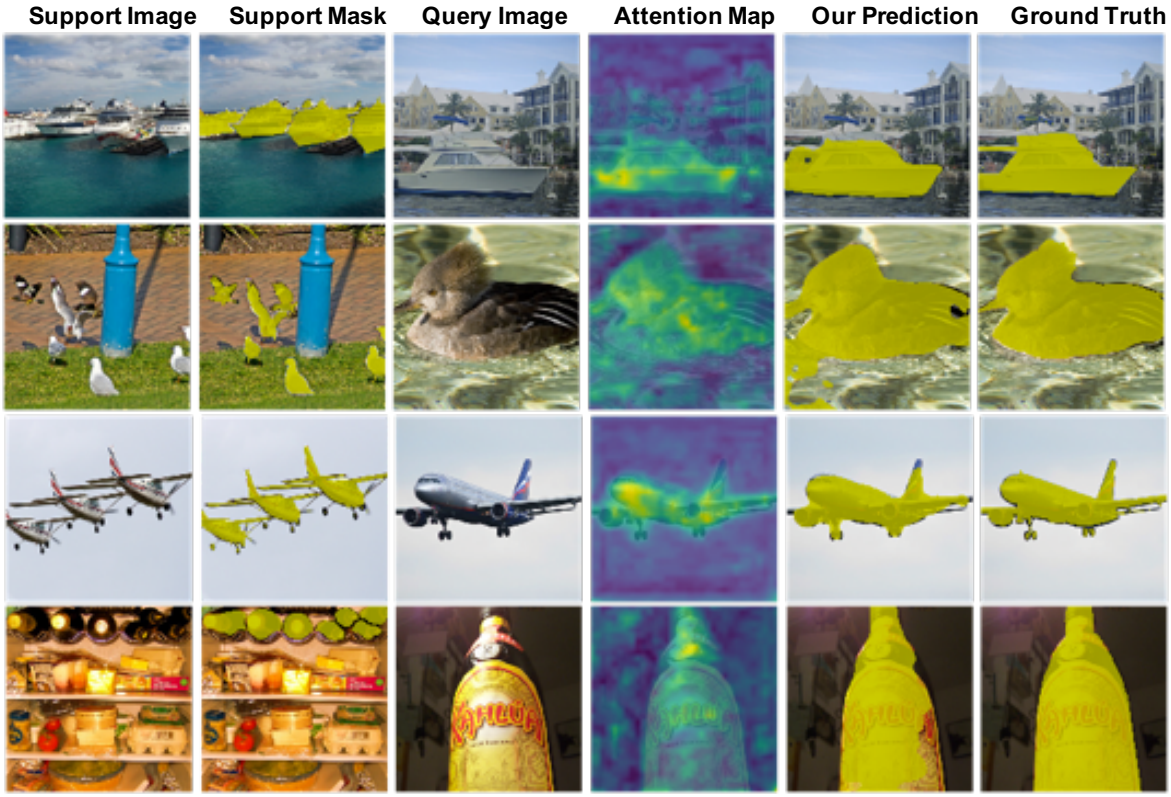

Figure 2. Example results under “many-to-one matching” scenario.

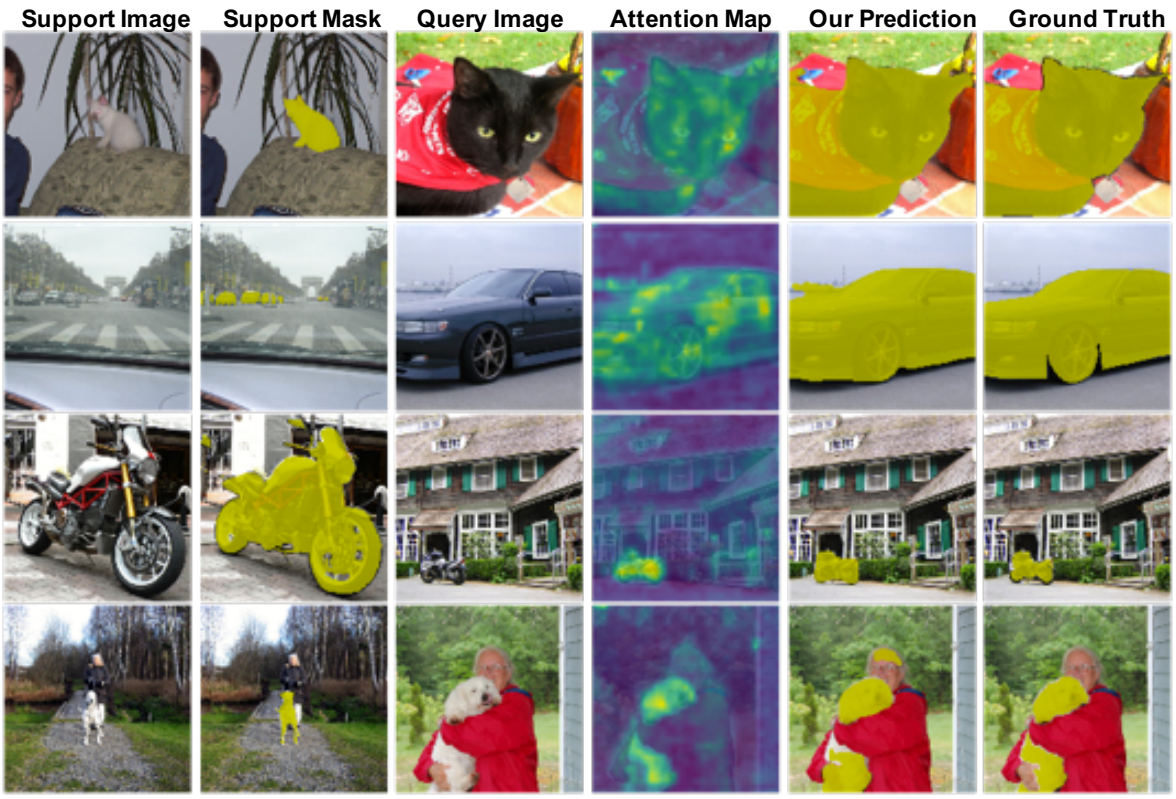

Figure 3. Example results when objects have large variations in object sizes.

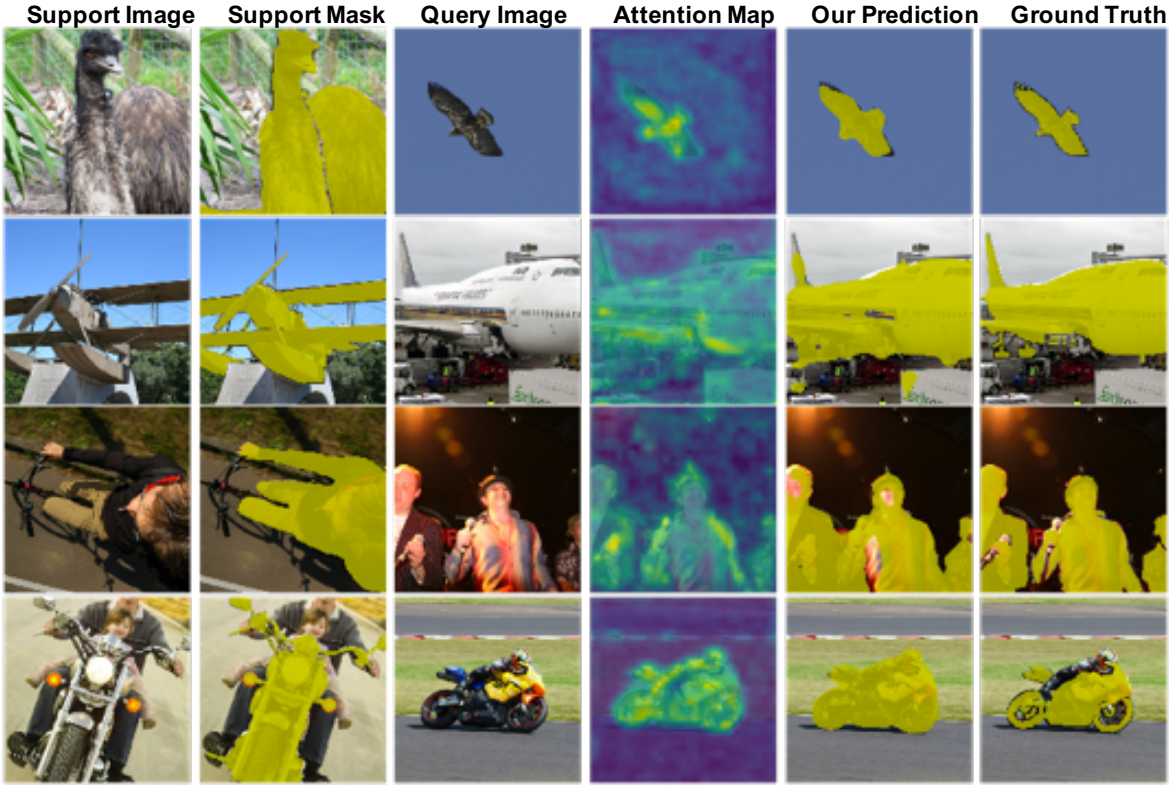

Figure 4. Example results when objects have large variations in viewing angles.

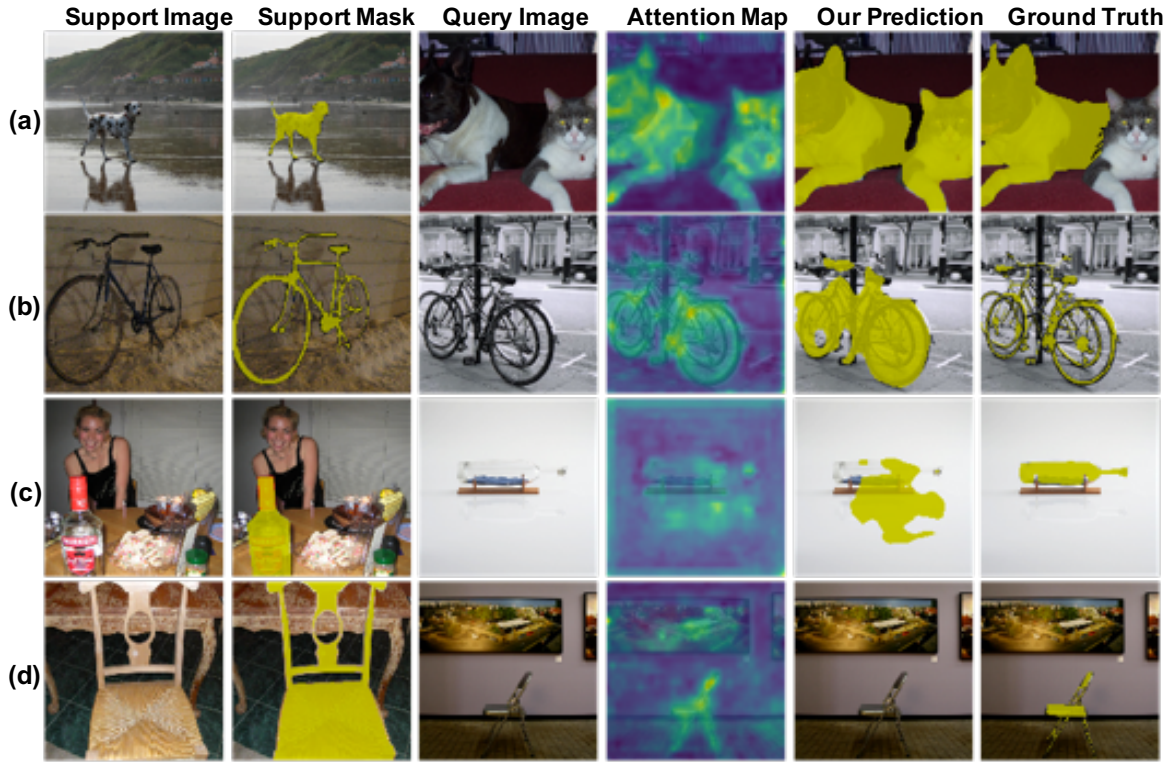

Figure 5. Selected failure cases. (a) The support example shows a dog while the query image has a dog and a cat. Our predicted mask covers both animals as they look very similar to each other. (b) The bicycle is difficult to segment due to its complicated structure, and our segmentation result is relatively coarse compared to the ground truth. However, our predicted mask does correctly cover the entire bicycle. (c) The query image in this example is very challenging as the bottle is transparent and has reflections on the surface which makes it difficult to perform the segmentation. Despite the poor segmentation results, we can still see that our attention map has activation on the region of the bottle and its reflection. (d) In this example, the chairs in the support image and query image are made of different materials and in different viewing angles. Although the final prediction does not correctly mask out this chair, we find that our attention map has high activation on the correct object regions.

Figure 6. More example results on Pascal-5^i dataset.

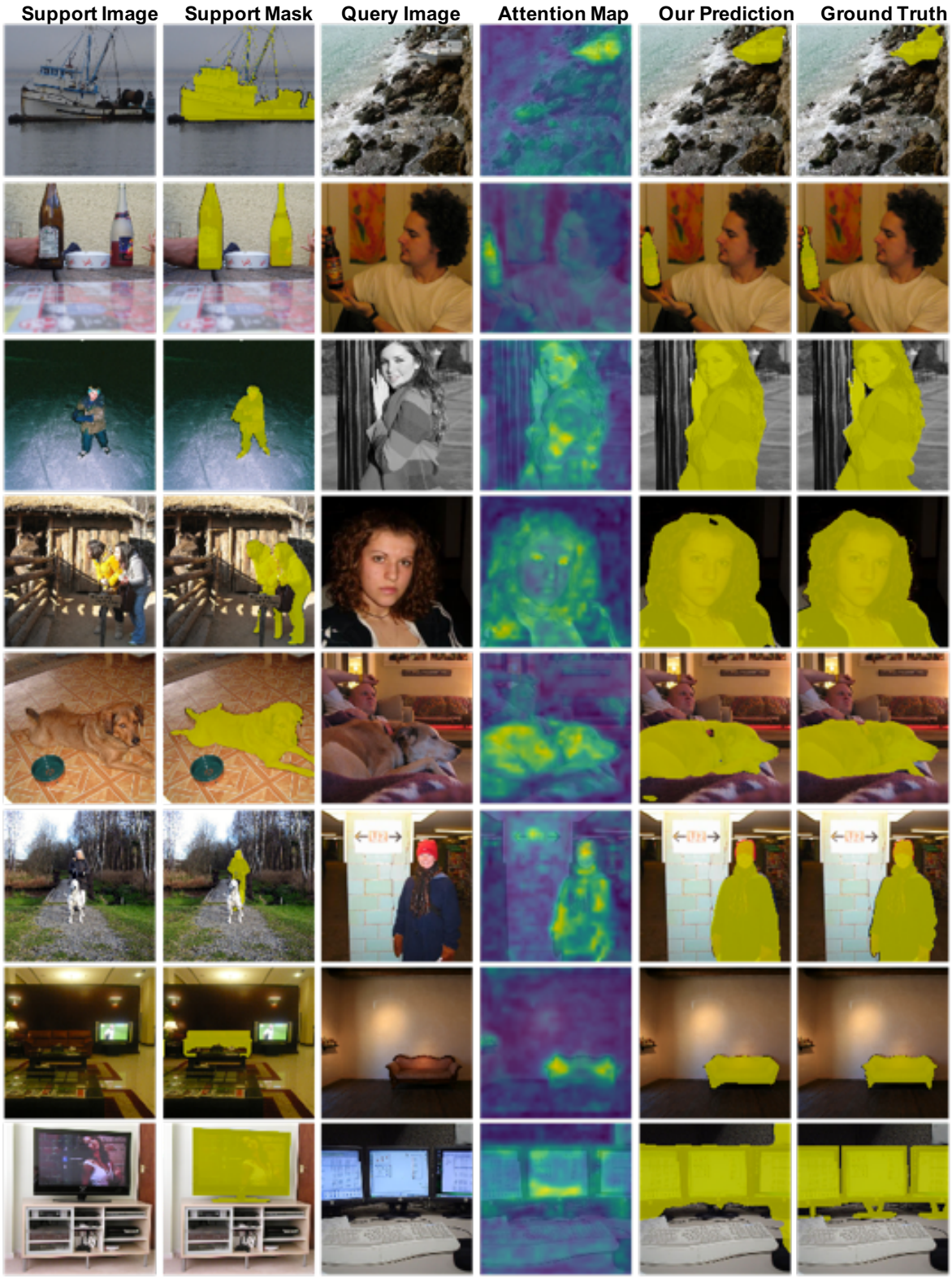

Figure 7. More example results on Pascal-5^i dataset.

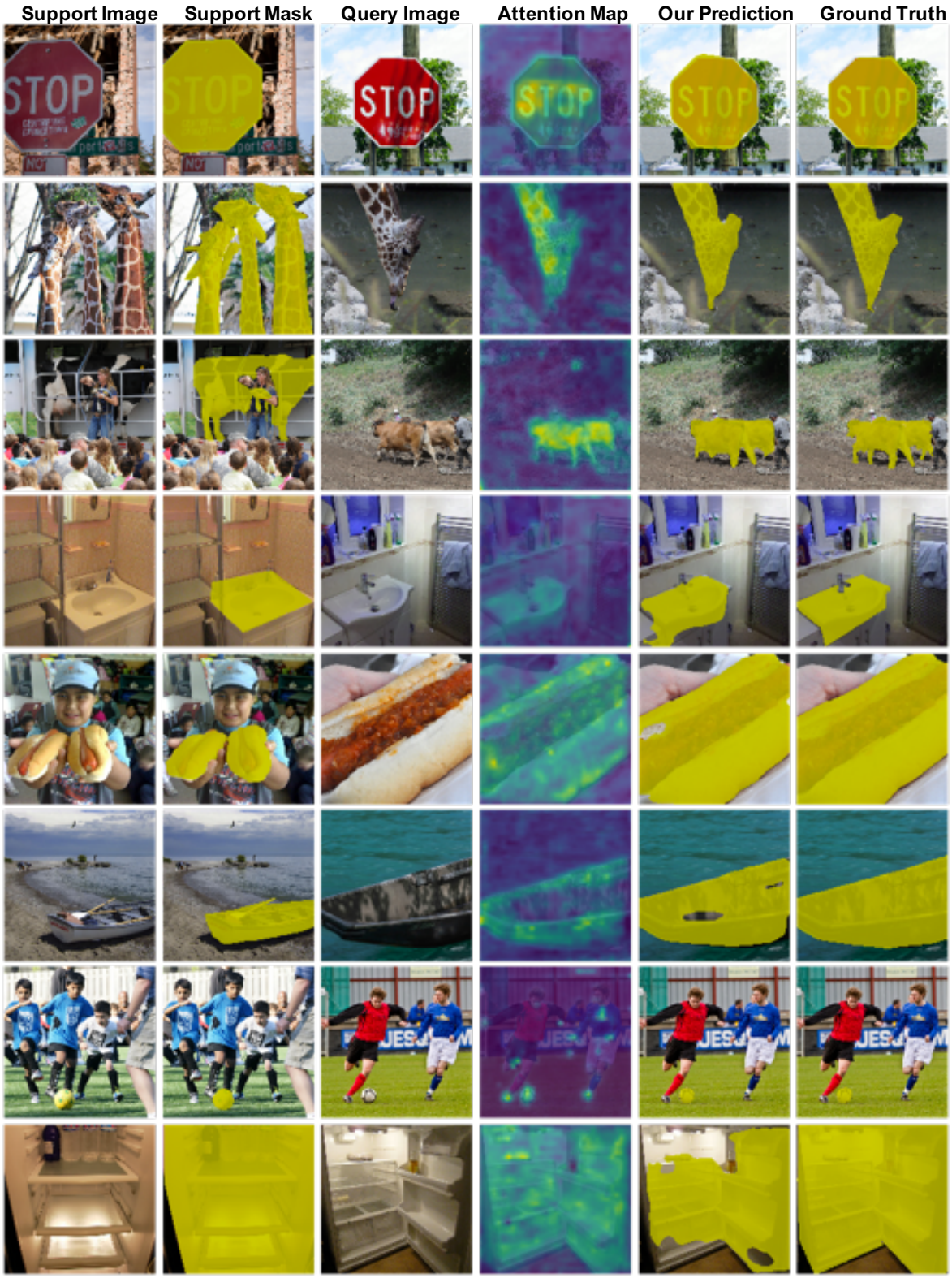

Figure 8. More example results on COCO-20^i dataset.

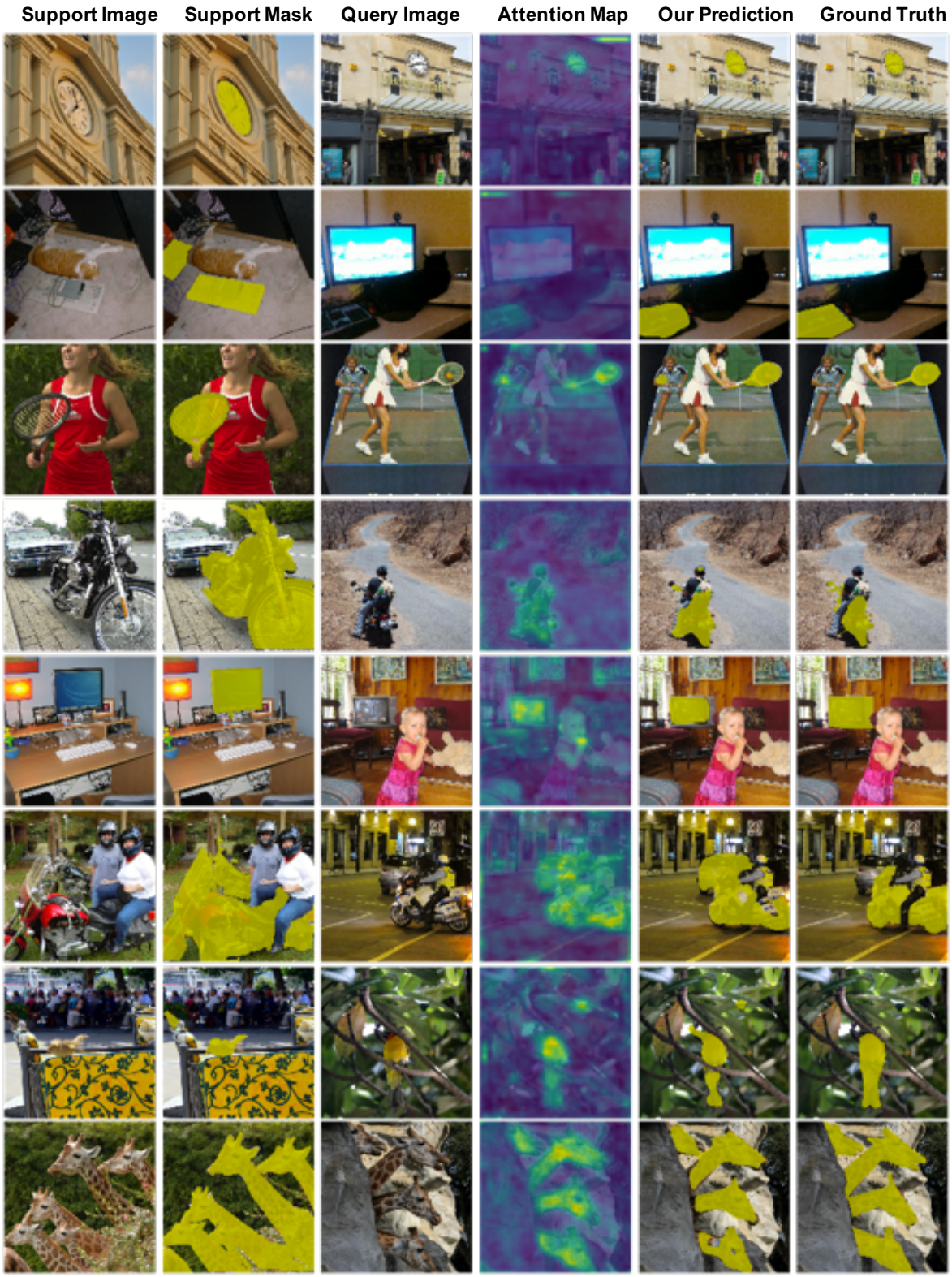

Figure 9. More example results on COCO-20^i dataset.