Supplementary Material

The inner mechanism of ConvLSTM.

ConvLSTM, a variant of Long Short Time Memory(LSTM), is originally proposed to predict the raindrops which is related to time series data.

In the meanwhile, there are some previous works which apply ConvLSTM in non-time series data. For example, ReferSeg(Li et al. 2018) utilized ConvLSTM to subsequently refine the feature map(Fig. 2 in (Li et al. 2018)) which has no time-series property as well. RIS(Romera-Paredes and Torr 2016) also used ConvLSTM to implement recurrent instance segmentation. There is also no time dependency in their input. **ConvLSTM's advantage is not only modeling the sequential data, but also sequentially filtering and fusing the data by gate mechanism in the ConvLSTM.**

For better analyzing the effect of ConvLSTM, we visualize the segmentation result by changing support image number from 1 to 5 as shown in Fig. 1. We can observe that the segmentation result tends to be close to ground truth as more support images are provided. On the contrary, the logical or fusion used in previous methods only realize a hard fusion so that the fusion result doesn't get obvious improvement. Specially, in the upper image example, when the second support image is fed into our network, the performance becomes unsatisfactory because the area of support image mask is too small. Interestingly, the segmentation performance is recovered by the later three support images for their explicit guiding.

In our paper, ConvLSTM can bring us two benefits: (1). arbitrary-shot learning for image segmentation by an end-to-end pattern. (2). dynamically fusing the previous support feature rather than meaningless logical or operation.

Visualization result

Some visualization result can be seen in Fig. 2. We also explore the effect of increasing the support set size as shown in Fig. 3.

References

Li, R.; Li, K.; Kuo, Y.-C.; Shu, M.; Qi, X.; Shen, X.; and Jia, J. 2018. Referring image segmentation via recurrent refinement networks. In *IEEE Conference on Computer Vision and Pattern Recognition*.

Romera-Paredes, B., and Torr, P. H. S. 2016. Recurrent instance segmentation. In *European Conference on Computer Vision*, 312–329. Springer.

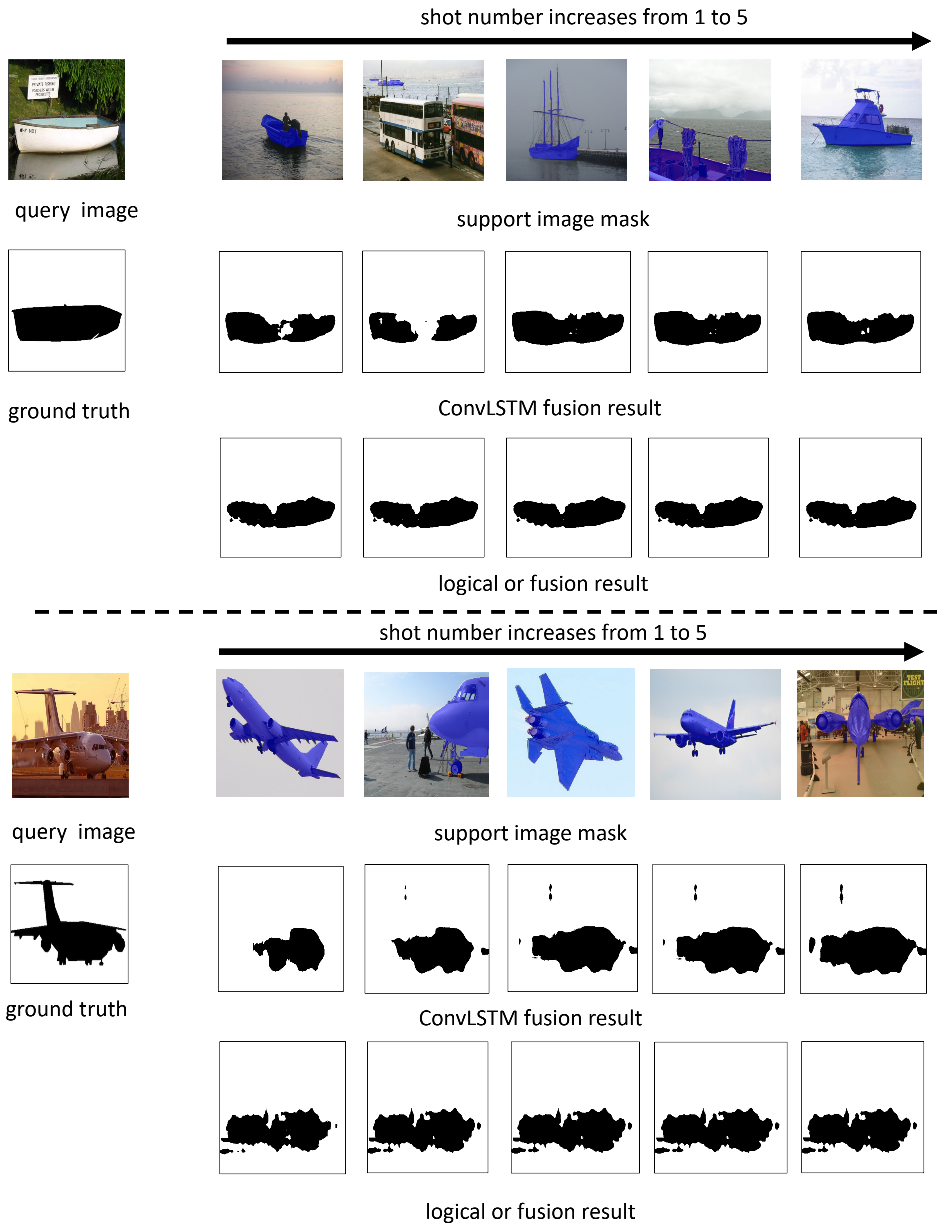

Figure 1: 5-shot segmentation result when feeding 5 support image masks gradually. ConvLSTM fusion result and the traditional logical or fusion result are mainly compared in this figure. Two image segmentation results are demonstrated. Best viewed in color.

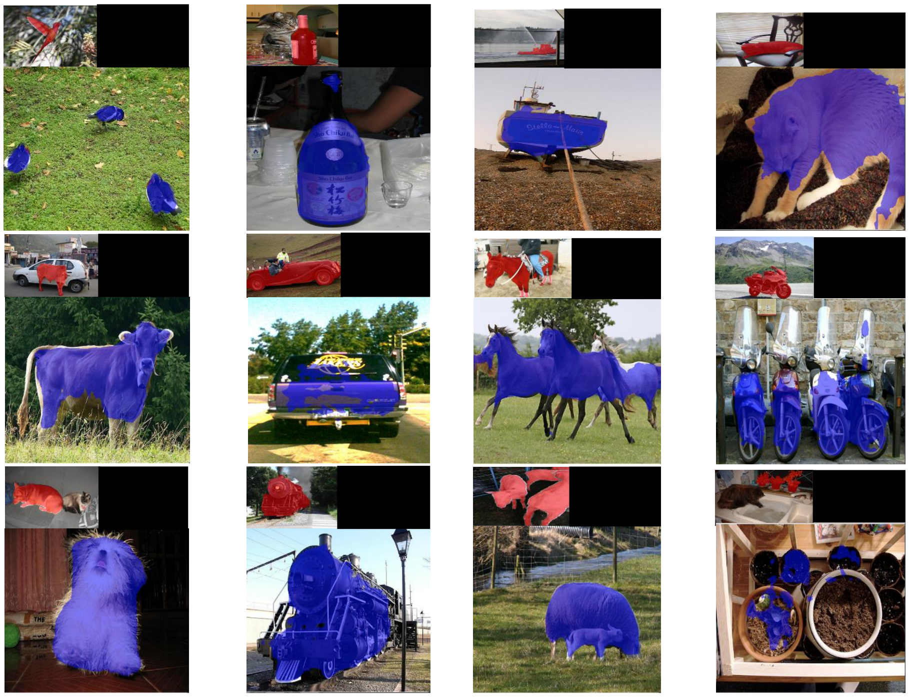

Figure 2: Some qualitative results of our method for 1-shot learning. Inside each tile, we have the support set at the top and the query image at the bottom. The support is overlaid with the ground truth in red and the query is overlaid with our predicted mask in blue. The last image shows the failure case. Best viewed in color.

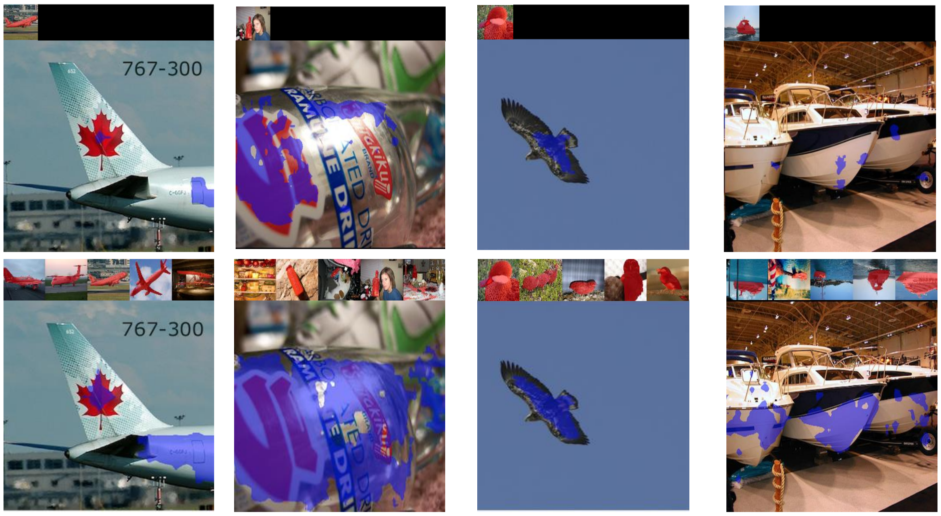

Figure 3: Effect of increasing the size of the support set. Results of 1-shot and 5-shot learning on the same query image are in the first and second rows respectively. The prediction is in blue. Best viewed in color.