The Devil is in the Boundary: Exploiting Boundary Representation for Basis-based Instance Segmentation
Supplementary Material

1. Implementation details

We verify our method mainly on the BlendMask [2] framework. Most hyperparameters are kept the same except the resolution of attention map 7×7 for the training efficiency. ResNet-50 [3] is used as our backbone network and the weights are pre-trained on ImageNet [4]. We train the model using stochastic gradient descent (SGD) with 4 TITAN RTX GPUs. We set the mini-batch size to 16 images. We adopt $1\times$ schedule (90K iterations) with the initial learning rate of 0.01. It is reduced by a factor of 10 at iteration $60K$ and $80K$, respectively. Input images are resized to have shorter side 800 pixels and a longer side at maximum 1333 pixels.

2. Qualitative Results

We provide additional qualitative results in Fig. 1, Fig. 2, Fig. 3, and Fig. 4. Fig. 1 shows the visual results of BlendMask [2] and ours. Fig. 2 shows the visual results of YOLACT [1] and ours. We can obviously observe that our method generates more accurate mask predictions. Moreover, we provide learned boundary basis on Fig. 3. Finally, the superior boundary segmentation results over the baseline BlendMask are in Fig. 4.

References

[1] Daniel Bolya, Chong Zhou, Fanyi Xiao, and Yong Jae Lee. Yolact: Real-time instance segmentation. In *Proc. of Int'l Conf. on Computer Vision (ICCV)*, 2019.

[2] Hao Chen, Kunyang Sun, Zhi Tian, Chunhua Shen, Yongming Huang, and Youliang Yan. Blendmask: Top-down meets bottom-up for instance segmentation. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, pages 8573–8581, 2020.

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, 2016.

[4] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Proc. of Neural Information Processing Systems (NIPS)*, 2012.

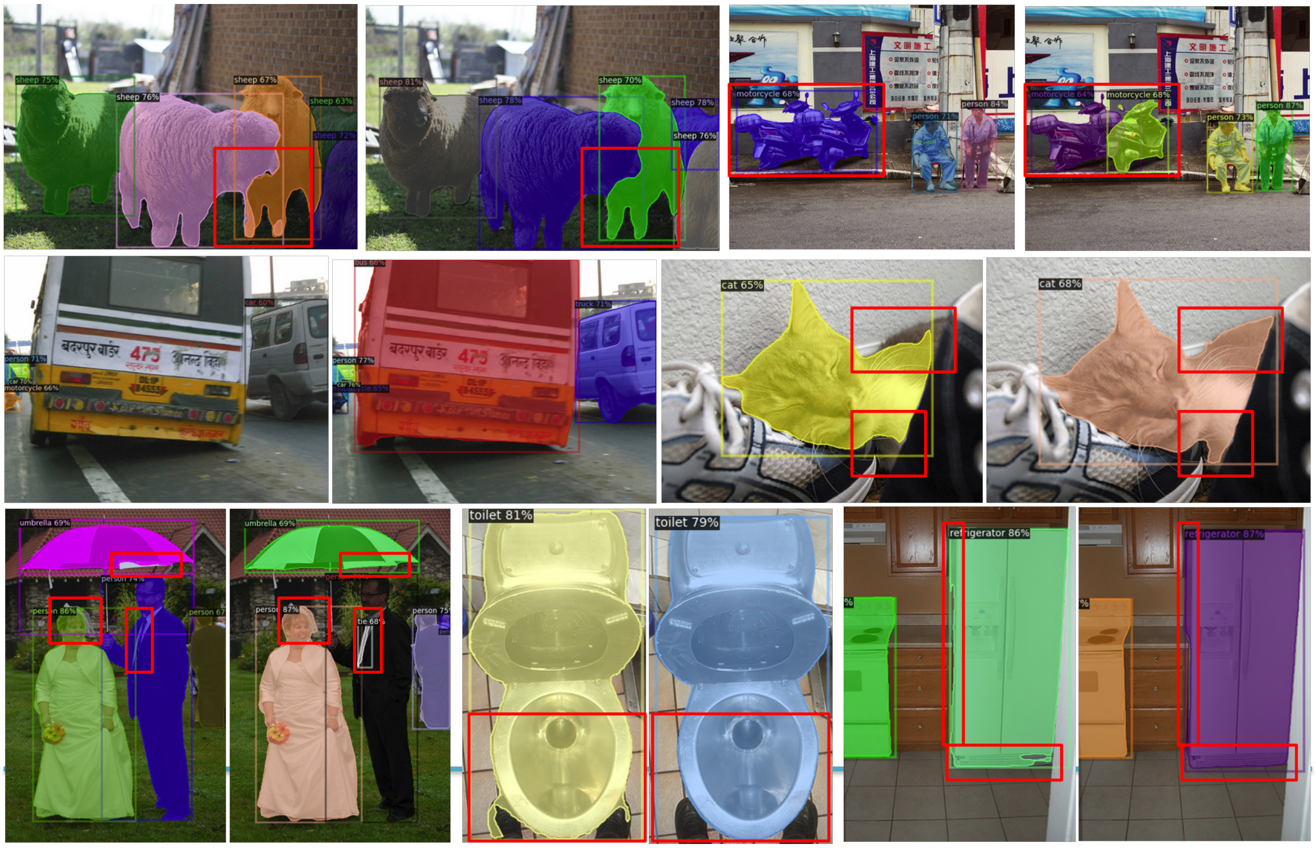

Figure 1. **Qualitative Comparisons between BlendMask [2] and *B2Inst*-BlendMask.** The image on the left is from the BlendMask while the image on the right is from ours.

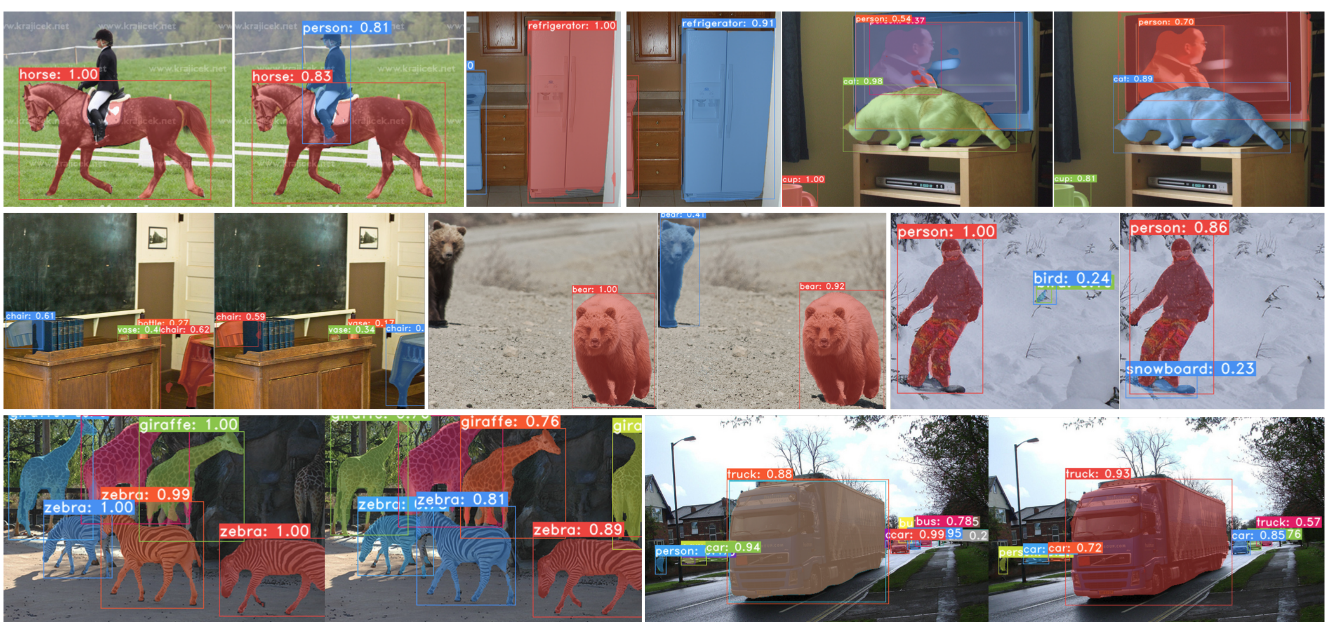

Figure 2. **Qualitative Comparisons between YOLACT [1] and *B2Inst*-YOLACT.** The image on the left is from the YOLACT while the image on the right is from ours.

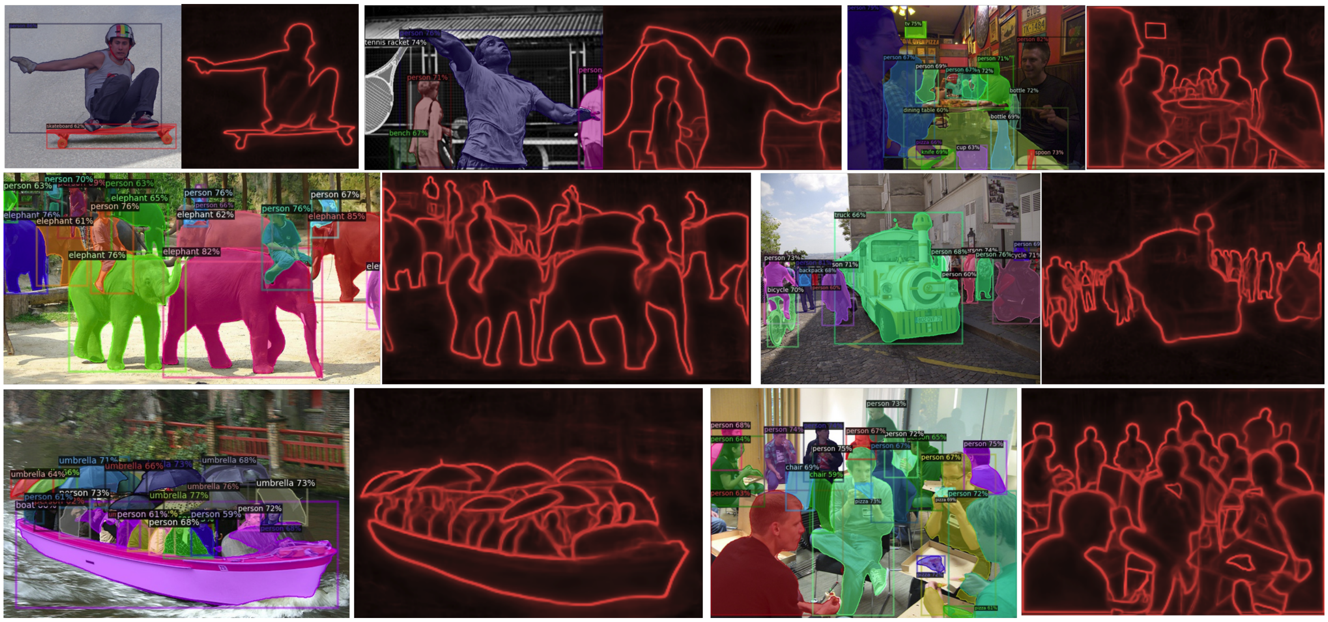

Figure 3. **The visualization of learned boundary basis.**

Figure 4. **Qualitative Comparisons on the Boundary Area.** (left) the results from BlendMask [2] (middle) the results from ours (right) the corresponding learned image boundary basis.