Deep Occlusion-Aware Instance Segmentation with Overlapping BiLayers (Supplemental material)

Lei Ke[1] Yu-Wing Tai[2] Chi-Keung Tang[1]

[1]The Hong Kong University of Science and Technology [2]Kuaishou Technology

{lkeab,cktang}@cse.ust.hk, yuwing@gmail.com

In this supplementary material, we provide more details about the proposed Synthetic Occlusion Dataset for instance segmentation, and more qualitative comparisons of the **modal** segmentation results on COCO [2] (Figure 4) and **amodal** segmentation results on KINS [3] (Figure 5).

1. Synthetic Occlusion Dataset

Complete Object Dataset We first collect a *complete object dataset* from COCO by conditionally filtering out the objects with bounding boxes overlapping rate over 5% and mask area smaller than 32×32, and then perform the complete object selection manually. This object dataset consists of images for *non-occluded single* object with corresponding complete mask and contour annotation. The collected dataset has 80 categories with instances number distribution as shown in Figure 1. Some random sample images of the complete objects are shown in Figure 2.

Occlusion Synthesis Process As shown in Figure 3, to diversify the occlusion patterns, we construct the large-scale **Synthetic Occlusion Dataset** by sampling both occluding and occluded instances from the complete object dataset following uniform class distribution. Then, a synthetic image based on the original image corresponding to the occluded target is produced by placing the occluding instance at a random image position (generated by grid search) which satisfies the object overlapping rate between 0.2 to 0.5. The synthetic occlusion dataset contains 100K such occluded images with **amodal** contours/masks for *both* occluding and partially occluded objects.

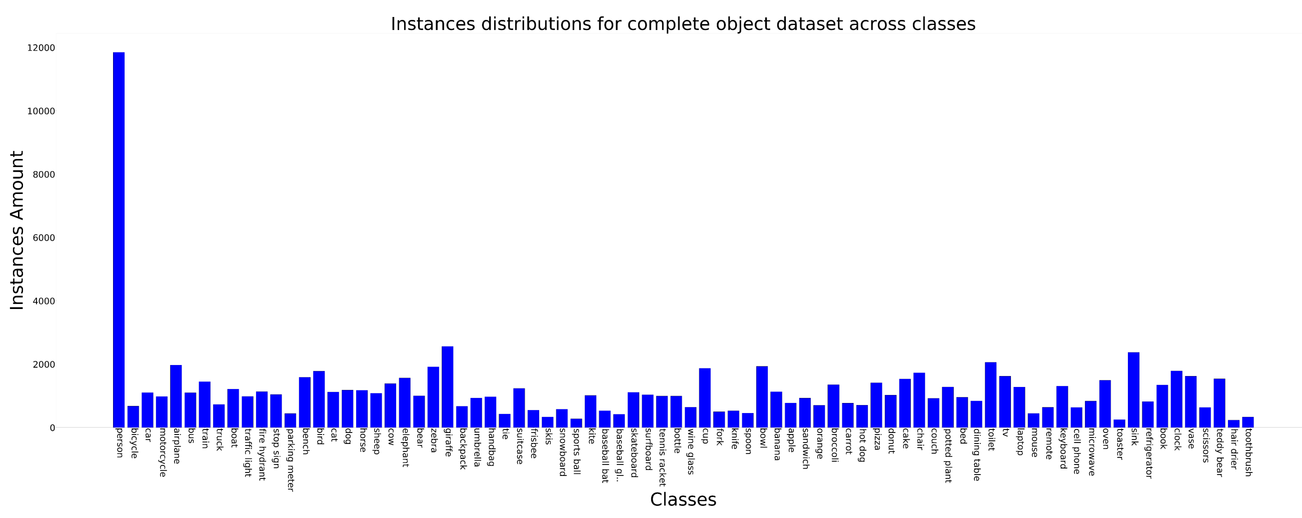

Figure 1. Instance distribution for the collected complete object dataset among 80 categories.

2. More Visualization Results

Modal results comparison on COCO More qualitative results of our proposed BCNet compared to the Mask Scoring R-CNN [1] on COCO *test-dev* set are shown in Figure 4, both using ResNet-101-FPN and Faster R-CNN detector [4]. In each

Figure 2. Sample object images from the collected Complete Object Dataset.

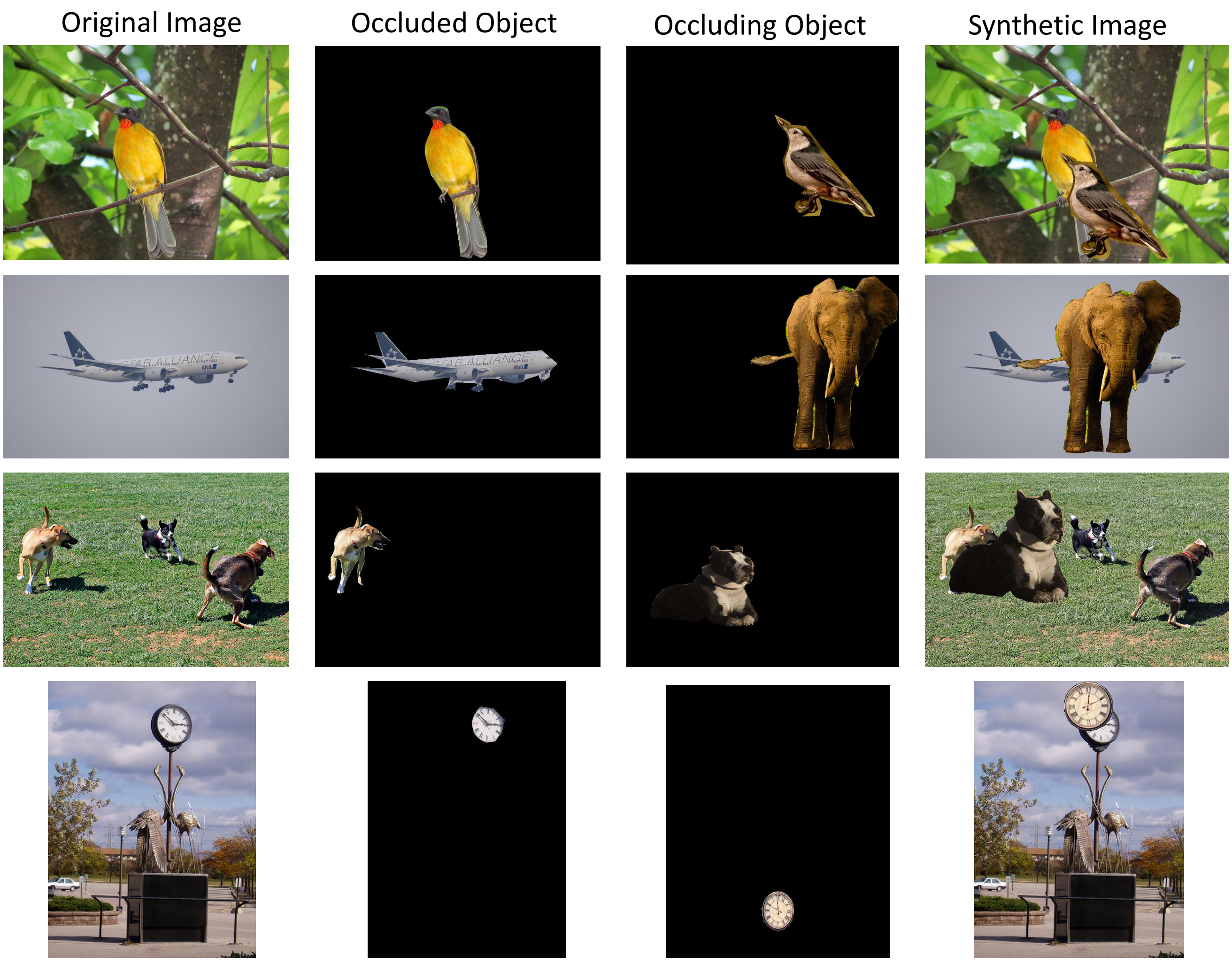

Figure 3. Occlusion synthesis process to produce the *Synthetic Occlusion Dataset* by sampling both occluding and occluded instances from the collected Complete Object Dataset, and grid searching the occluded positions in the image.

ROI region, GCN-1 detects occluding regions while GCN-2 models the partially occluded instance by directly regressing the contours and masks. Our proposed method is robust enough to deal with various occlusion cases, such as highly overlapping

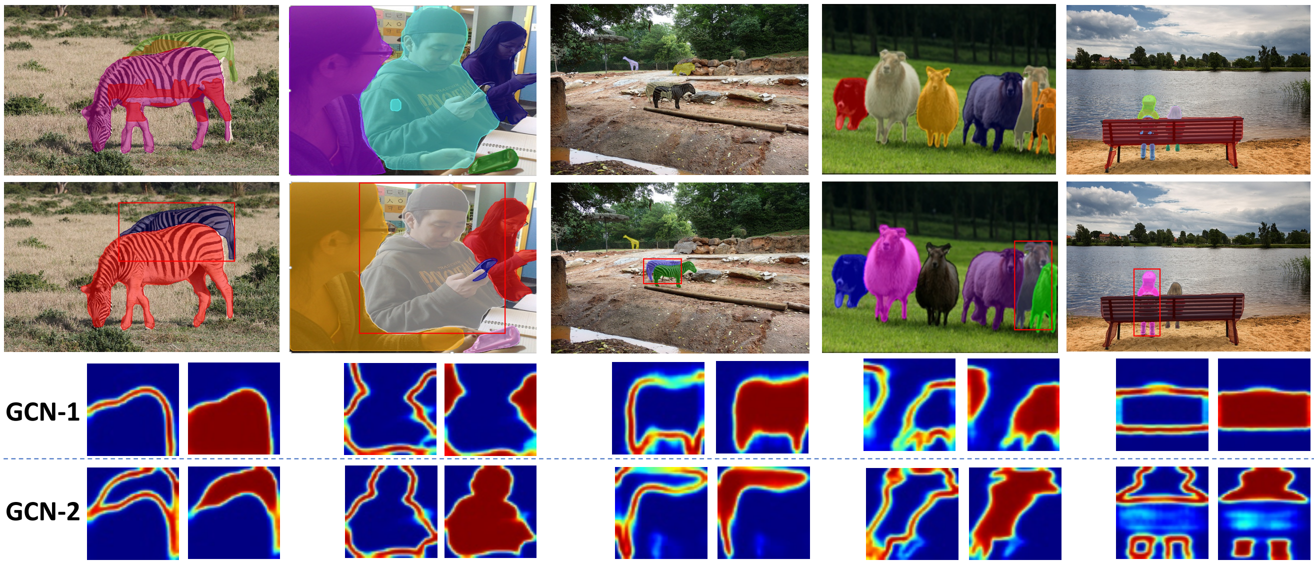

Figure 4. Qualitative **modal** results of Mask Scoring R-CNN [1] (top row) and our BCNet (middle row) on **COCO** *test-dev* set, both using ResNet-101-FPN and Faster R-CNN. The bottom row visualizes squared heatmap of contour and mask predictions by the two GCN layers for the occluder and occludee in the same **ROI region** specified by the red bounding box, which also makes the final segmentation result of BCNet more explainable than previous methods.

zebras and human hands. Also, the contour and mask predictions by the two GCN layers for the occluder (GCN-1) and occludee (GCN-2) in the same ROI region also makes the results of BCNet more explainable compared to previous methods.

Amodal results comparison on KINS In Figure 5, we additionally provide qualitative **amodal** segmentation results comparison between Mask R-CNN + ASN module [3] and our BCNet on KINS [3] test set. Take the first case as an example, our BCNet infers more reasonable amodal car shape even when the front part of the car is heavily occluded by the standing woman.

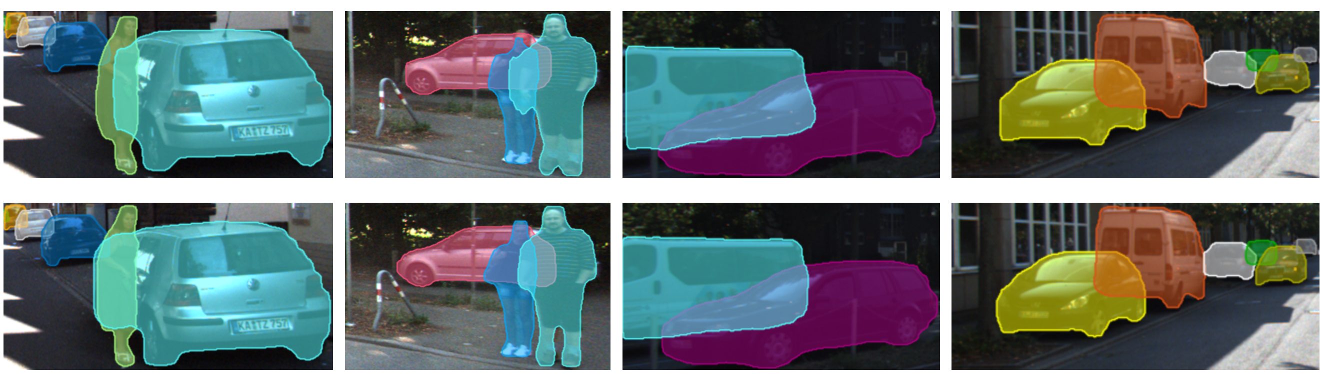

Figure 5. Additional qualitative **amodal** results comparison between Mask R-CNN + ASN module [3] (top row) and our BCNet (bottom row) for the mask predictions on **KINS** test set [3], both using ResNet-101-FPN and and Faster R-CNN detector [4], where the mask shape of the **invisible/occluded regions** are more reasonably estimated by BCNet.

References

[1] Zhaojin Huang, Lichao Huang, Yongchao Gong, Chang Huang, and Xinggang Wang. Mask scoring r-cnn. In *CVPR*, 2019. 1, 3

[2] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *ECCV*, 2014. 1

[3] Lu Qi, Li Jiang, Shu Liu, Xiaoyong Shen, and Jiaya Jia. Amodal instance segmentation with kins dataset. In *CVPR*, 2019. 1, 3

[4] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *NeurIPS*, 2015. 1, 3