Supplementary Material for Conditional Convolutions for Instance Segmentation

Zhi Tian, Chunhua Shen, Hao Chen

The University of Adelaide, Australia

1 Experiments on Cityscapes

We also conduct experiments on the benchmark Cityscapes [1]. The benchmark has fine annotations for 2, 975 training, 500 validation, and 1, 525 testing images. Following Mask R-CNN [2], we only use the fine annotations to train the models. The results are reported on the Cityscapes `val` split. The training details are the same as that of Mask R-CNN on Cityscapes in `Detectron2` [3]. As shown in Table 1, CondInst outperforms Mask R-CNN by 0.4% AP on Cityscapes `val`.

Table 1: Results on Cityscapes `val` with ResNet-50-FPN.

method	AP	AP_{50}	person	rider	car	truck	bus	train	mcycle	bicycle
Mask R-CNN	36.5	62.2	**35.0**	**28.8**	53.0	34.3	57.0	**37.5**	**22.3**	**23.8**
CondInst	**36.9**	**63.2**	**35.0**	28.4	**55.5**	**37.4**	**57.5**	36.3	22.1	23.3

2 More Visualization Results

Here, we show more qualitative results in Fig. 1.

References

1. Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B.: The cityscapes dataset for semantic urban scene understanding. In: Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
2. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask R-CNN. In: Proc. IEEE Int. Conf. Comp. Vis. pp. 2961–2969 (2017)
3. Wu, Y., Kirillov, A., Massa, F., Lo, W.Y., Girshick, R.: Detectron2. `https://github.com/facebookresearch/detectron2` (2019)

Fig. 1: More qualitative results of CondInst.