Supplemental Materials to
" DynaMask: Dynamic Mask Selection for Instance Segmentation"

In this supplemental file, we provide the following materials:

- Analyses on mask resolution prediction (*cf*. Sec4.3-Mask Resolution Prediction in the main paper);
- Distribution of predicted mask resolutions (*cf*. Sec4.3-Size-based Mask Selection in the main paper);
- Results on LVIS dataset;
- Comparisons with other variants of FPN;
- Qualitative results (*cf*. Sec4.2-Visualizations of Segmentation Results in the main paper).

A. Analyses on Mask Resolution Prediction

Correlation between mask resolution and class. We show the mask distributions of different classes for DynaMask ($54\% \downarrow$) in Fig. S1. As can be seen, the classes with irregular shapes (*e.g.*, "giraffe") tend to be assigned larger masks, while the classes of regular shapes (*e.g.*, "book") are allocated smaller masks for efficiency.

Mask selection results. Fig. S2 shows more examples of mask resolution selection. As can be seen, the "hard" objects with irregular shapes and complex boundaries are assigned larger masks, such as the "potted plant", "person", *etc*. On the contrary, the "easy" samples with regular shapes and less details are assigned smaller masks, such as the "skis", "suitcase", "dining table" and so on.

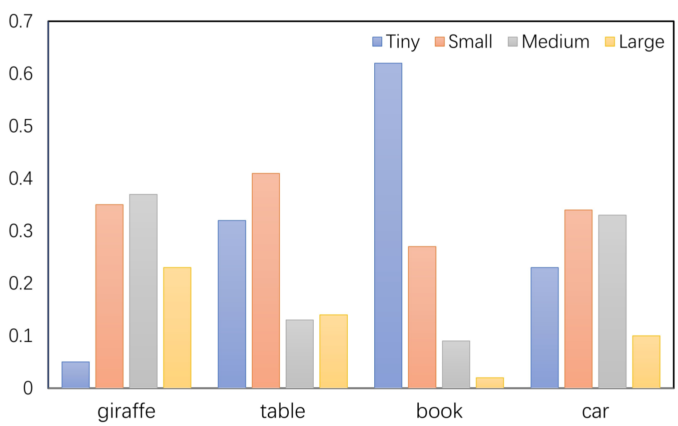

Figure S1. The mask distributions of different classes.

B. Distribution of Predicted Mask Resolutions

To better understand why our method can reduce the computations, we count the number of masks assigned to different resolutions and report them in Tab. S1. We take "Size-Based" approach, which assigns mask according to the size of objects (please refer to Eq. 8 in the main paper), as a reference here. One can see that by selecting lower-resolution masks, the

FLOPs can be reduced by 19% without performance degradation, demonstrating that it is redundant to use large masks for all instances. DynaMask (54% $\downarrow$) outperforms Size-Based approach by 0.9% AP with comparable FLOPs, indicating that the proposed dynamic mask selection is more effective to select suitable masks.

Method	Tiny	Small	Med	Large	FLOPs	AP
Size-Based	0.47	0.32	0.14	0.08	0.56G	35.9
DynaMask	0	0	0	1	1.4G	37.6
DynaMask (19% $\downarrow$)	0.05	0.13	0.25	0.56	1.13G	37.6
DynaMask (54% $\downarrow$)	0.35	0.34	0.21	0.10	0.64G	36.8

Table S1. The distribution of predicted masks.

C. Results on LVIS Dataset

Since the quality of mask annotations in COCO [6] is limited, which makes it a less convincing candidate for validating the accuracy of high-resolution mask prediction. We hence evaluate the performance on LVIS [3], which has higher-quality annotations. The corresponding results are reported in Tab. S2 below. DynaMask outperforms Mask R-CNN [4] by 3.6% AP since we apply higher-resolution masks to "harder" samples, which usually contain more details.

Method	Backbone	AP	AP_r	AP_c	AP_f
Mask R-CNN [4]	R50-FPN	22.1	10.1	21.7	31.7
DynaMask	R50-FPN	25.7	13.9	24.4	32.0

Table S2. Results on LVIS validation dataset.

D. Comparisons with Other Variants of FPN

We adopt the Dual-FPN to integrate region-level features for refining mask predictions. We compare our method with other variants of FPN, such as PANet [7], BiFPN [2], and NAS-FPN [8]. The results are reported in Tab. S3. One can see that our method performs better since we adaptively aggregate complementary information from different feature levels.

method	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
PANet [7]	35.9	56.8	38.3	19.4	39.2	49.0
BiFPN [2]	36.4	57.2	39.0	19.2	39.8	49.5
NAS-FPN [8]	36.0	56.9	38.7	18.9	38.5	49.9
DynaMask (ours)	37.6	57.4	40.5	20.7	40.4	50.3

Table S3. Comparisons with different variations of FPN.

E. Qualitative Results

In Figs. S3-S5, we visualize the segmentation results of DynaMask together with two representative methods: Mask R-CNN [4] and PointRend [4]. As can be seen, our method achieves finer and more accurate predictions, especially on the boundary regions of objects.

References

[1] Tianheng Cheng, Xinggang Wang, Lichao Huang, and Wenyu Liu. Boundary-preserving mask r-cnn. In *European conference on computer vision*, pages 660–676. Springer, 2020.

[2] Golnaz Ghiasi, Tsung-Yi Lin, and Quoc V Le. Nas-fpn: Learning scalable feature pyramid architecture for object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7036–7045, 2019. 2

[3] Agrim Gupta, Piotr Dollar, and Ross Girshick. Lvis: A dataset for large vocabulary instance segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5356–5364, 2019. 2

[4] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017. 2, 5, 6, 7

[5] Alexander Kirillov, Yuxin Wu, Kaiming He, and Ross Girshick. Pointrend: Image segmentation as rendering. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9799–9808, 2020. 5, 6, 7

[6] 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 *European conference on computer vision*, pages 740–755. Springer, 2014. 2

[7] Shu Liu, Lu Qi, Haifang Qin, Jianping Shi, and Jiaya Jia. Path aggregation network for instance segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8759–8768, 2018. 2

[8] Mingxing Tan, Ruoming Pang, and Quoc V Le. Efficientdet: Scalable and efficient object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10781–10790, 2020. 2

Figure S2. Mask selection results by DynaMask. Each color corresponds to a candidate resolution (red→large, blue→medium, green→small, yellow→tiny).

Mask R-CNN **PointRend** **DynaMask**

Figure S3. Qualitative comparisons between Mask R-CNN [4], PointRend [5], and DynaMask, using ResNet-50-FPN backbone.

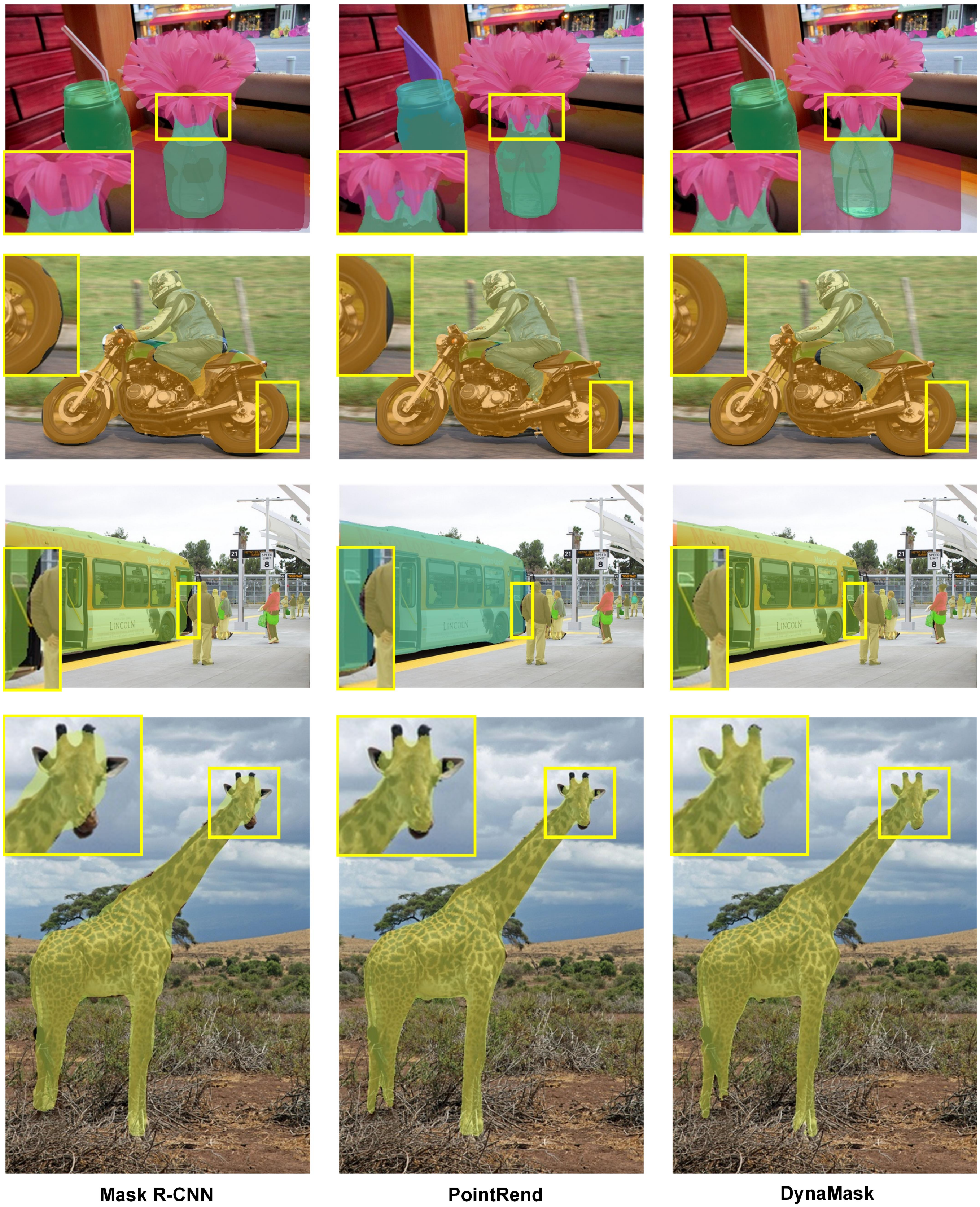

Figure S4. Qualitative comparisons between Mask R-CNN [4], PointRend [5], and DynaMask, using ResNet-50-FPN backbone.

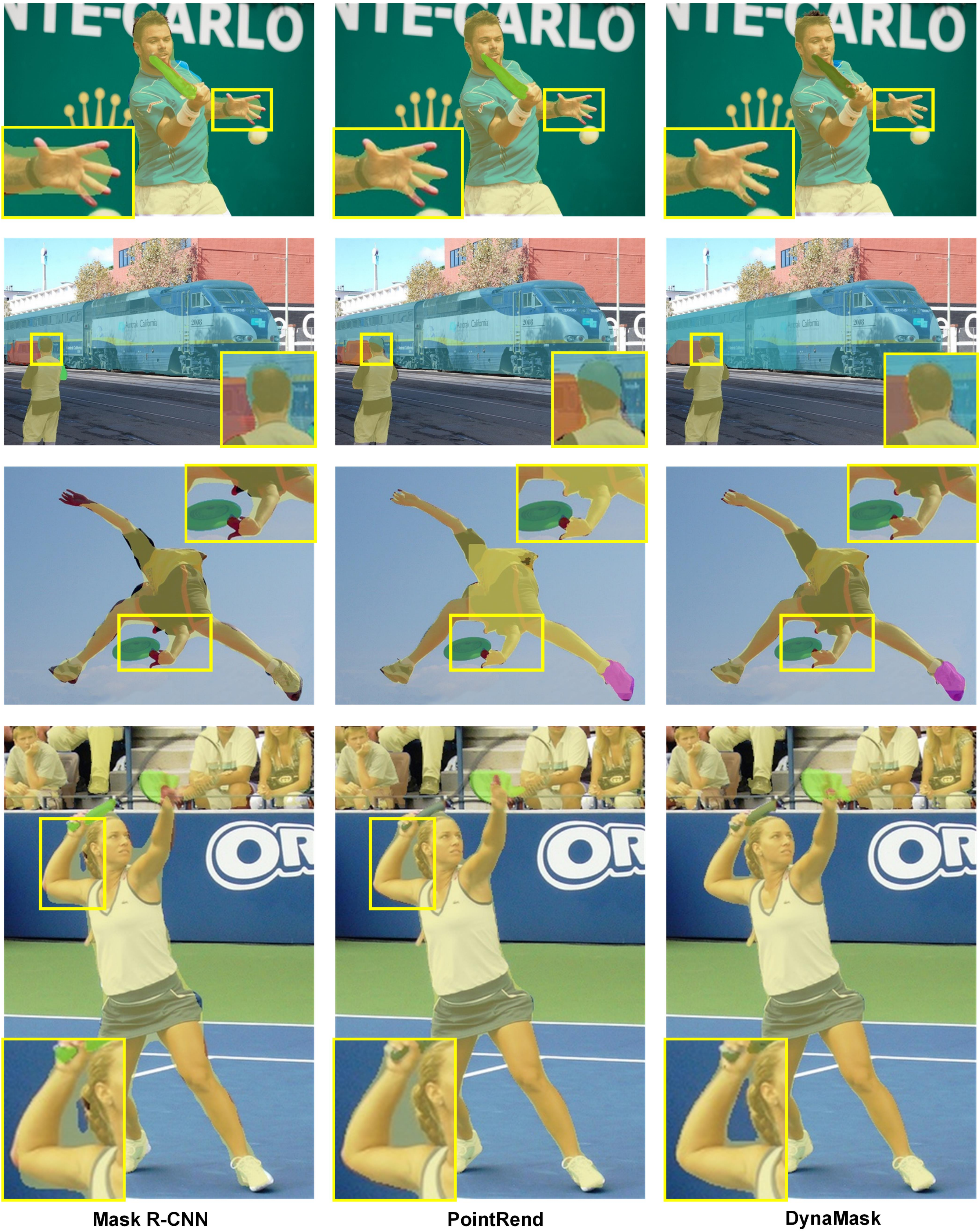

Figure S5. Qualitative comparisons between Mask R-CNN [4], PointRend [5], and DynaMask, using ResNet-50-FPN backbone.