Supplementary Material for A Simple-but-effective Baseline for Training-free Class-Agnostic Counting

Yuhao Lin*, Haiming Xu*, Lingqiao Liu, Javen Qinfeng Shi
Australian Institution for Machine Learning
The University of Adelaide
yuhao.lin01,hai-ming.xu,lingqiao.liu,javen.shi@adelaide.edu.au

In this supplementary material, we provide more details to complement the manuscript, including threshold sensitivity, visualization of generated mask proposal and qualitative results on CARPK dataset.

1. Threshold Sensitivity

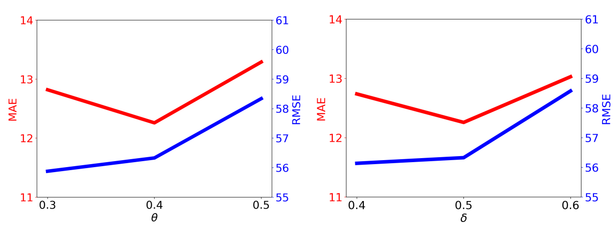

Figure 1. Performance with various threshold θ and δ.

Our method involves two hyperparameters: a similarity threshold θ for proposal selection and δ for transductive prototype updating. Figure 1 illustrates the robustness of our method to variations in the hyperparameters.

2. Comparison with PseCo

Both methods leverage foundation models for CAC, using SAM for object segmentation and semantic-rich encoders for final counting decisions. Key differences include: **Approach:** Our method demonstrates foundation model capabilities without additional trained modules, while PseCo incorporates two extra trained components (point decoder and object classifier). **Performance:**Our method outperforms PseCo on FSC-147, especially in RMSE (56.33 vs 112.86). This improvement is notable in images with crowded objects. As shown in Fig. 2, when counting Lego block protrusions (2560 total), PseCo's inaccurate point decoder leads to poor SAM segmentation and its object classifier predicts a count of 0. In contrast, our method, using superpixels for object-prior point prompts and a prototype-based classifier, identifies 2054 objects.

*Equally contributed.

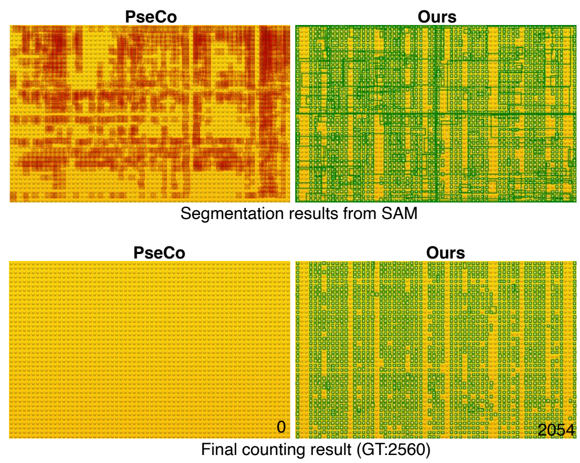

Figure 2. FSC-147 example results. Zoom in for detail.

3. Visualization of Generated Mask Proposal

In the main paper, we have presented the effectiveness of the use of superpixel on the quality of mask proposals generated by SAM in Figure 3. In this section, we further illustrate the visualization of generated mask proposals in Figure 3. It clearly shows that SAM with superpixel can achieve a high recall rate for the interested object without demanding of denser grid of points. Meanwhile, we find that SAM with superpixel can also generate mask proposals for very thin objects thanks to the object-prior prompt, such as the street lamp marked in the red box. Nonetheless, employing SAM with a standard grid of points might not consistently achieve segmentation for such objects.

4. Qualitative Results on CARPK

In this section, we further provide some qualitative visualization for the CARPK dataset. As shown in Figure 4, our method can generate precise masks for each car instance and achieve accurate counting results.

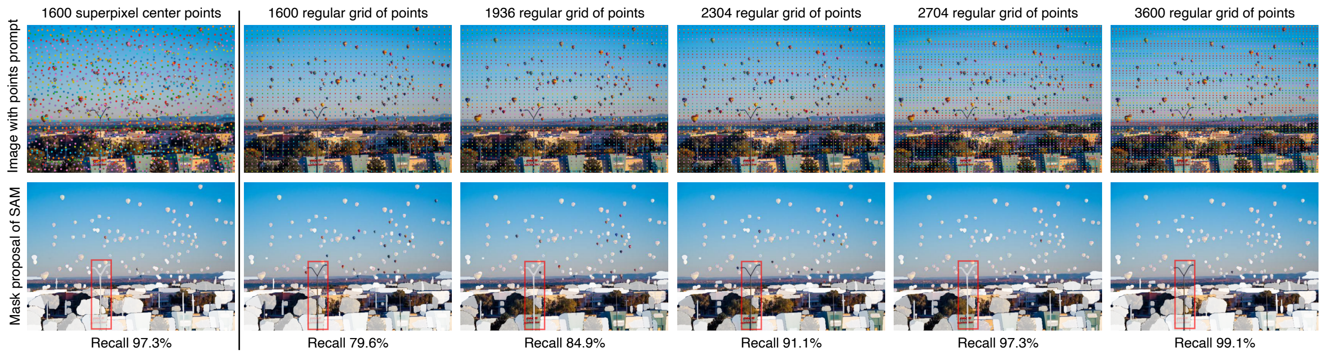

Figure 3. Visualization of mask proposals generated by SAM with various point prompts. In this example, the hot-air balloon is the user-interested object and the bottom row presents the recall of this object. The red boxes display mask generation for a very fine object (the street lamp).

Figure 4. Qualitative results on the CARPK dataset are displayed. Counting values are noted in the bottom-left corner. Best viewed by zooming in.