Supplementary Material for "Full-scale Selective Transformer for Semantic Segmentation"

Fangjian Lin[1,2,3,*], Sitong Wu[2,*], Yizhe Ma[1], and Shengwei Tian[1**]

[1] School of Software, Xinjiang University, Urumqi, China
[2] Baidu VIS, Beijing, China
[3] Institute of Deep Learning, Baidu Research, Beijing, China
wusitong98@gmail.com, {linfangjian01, mayizhe01, tianshengwei}@163.com

1 Visualizations

For better understanding our method, we visualize feature selection of query features. Examples from PASCAL Context [3], ADE20K [4], COCO-Stuff 10K [1], and Cityscapes[2] are shown in Figure 1, 2, 3, and 4, respectively. The i-th column shows the feature selection of query scale S_i. From these results, one can be seen that high-level semantic features tend to select low-level features with detailed spatial information, and vice versa.

* Equal contributions
** Corresponding author

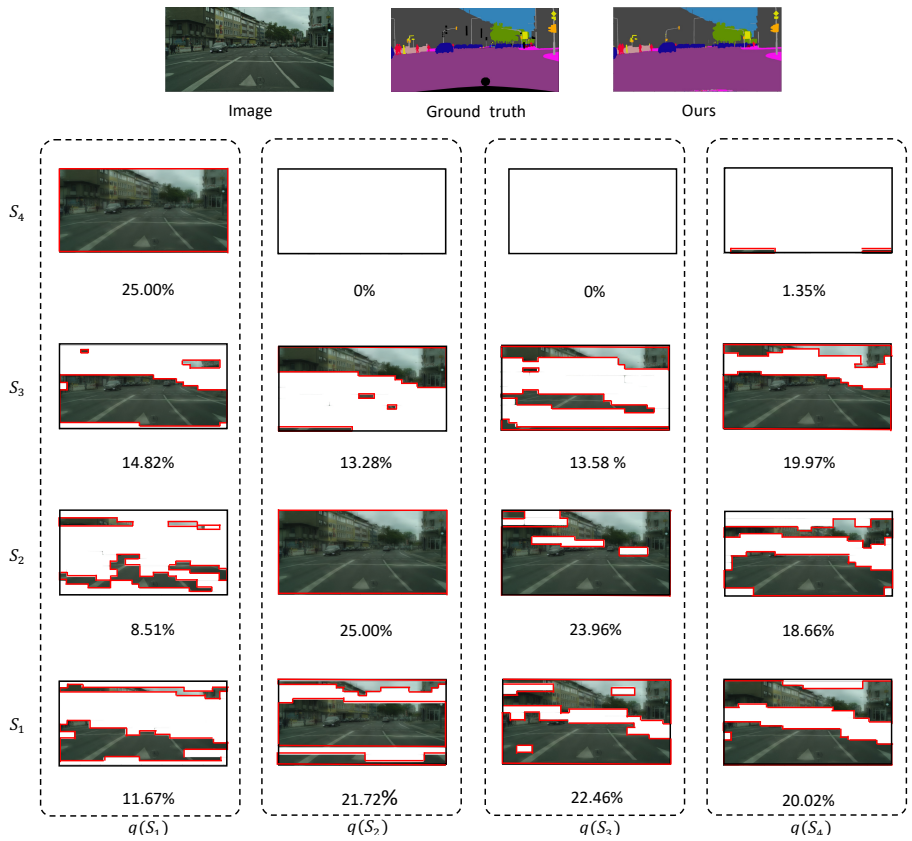

Fig. 1: Visualization of multi-scale feature selection on Cityscapes dataset. $q(S_i)$ indicates taking features from the stage or scale S_i as a query. The i-th column shows the feature selection of query scale S_i. The red polygon represents the selection area.

Fig. 2: Visualization of multi-scale feature selection on COCO-Stuff 10K dataset. $q(S_i)$ indicates taking features from the stage or scale S_i as a query. The i-th column shows the feature selection of query scale S_i. The red polygon represents the selection area.

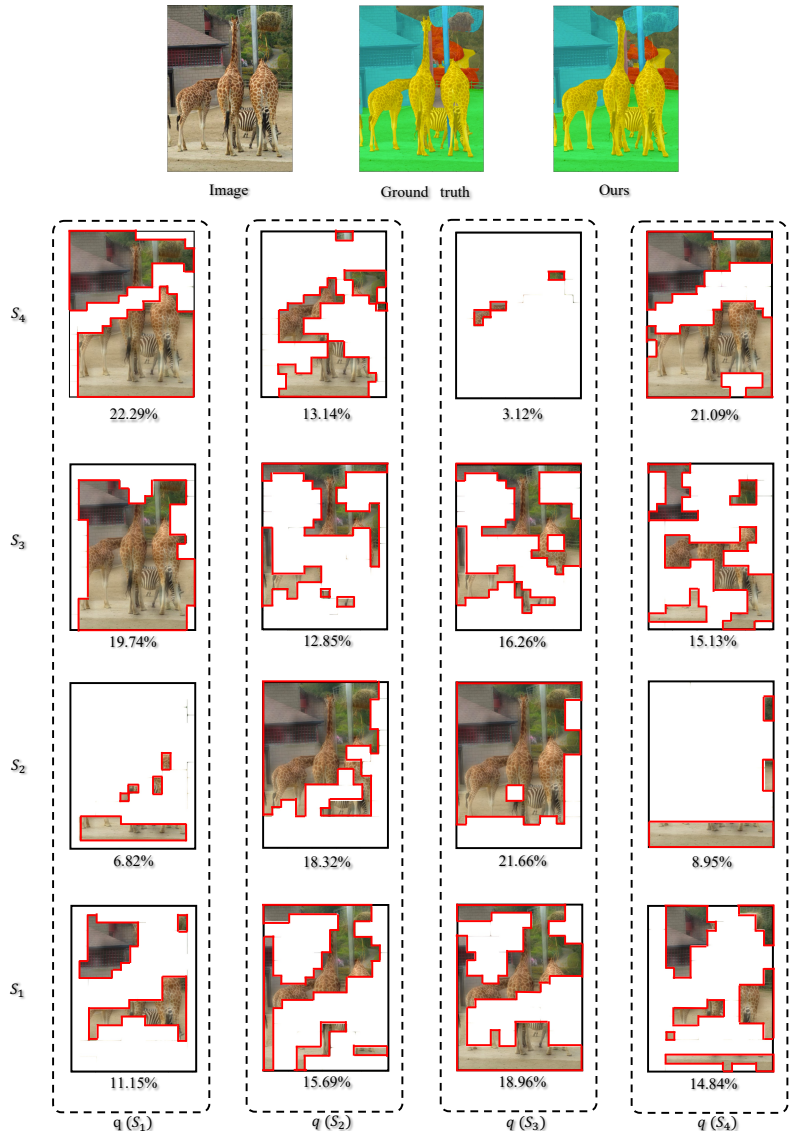

Fig. 3: Visualization of multi-scale feature selection on ADE20K dataset. $q(S_i)$ indicates taking features from the stage or scale S_i as a query. The i-th column shows the feature selection of query scale S_i. The red polygon represents the selection area.

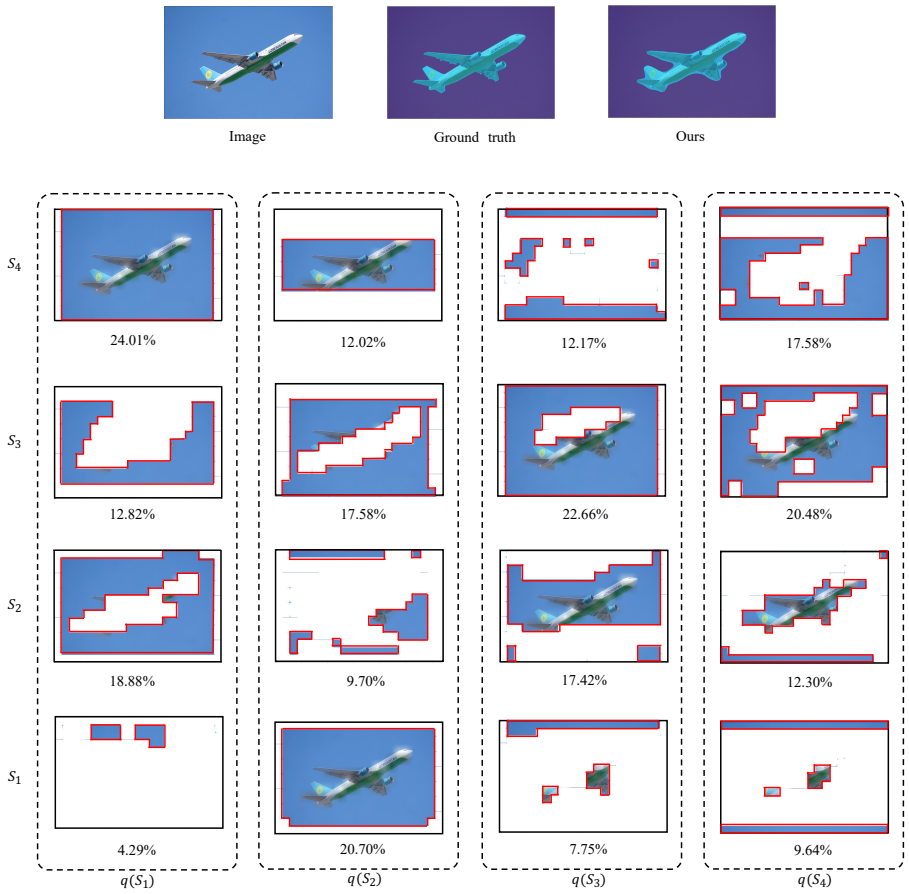

Fig. 4: Visualization of multi-scale feature selection on PASCAL Context dataset. $q(S_i)$ indicates taking features from the stage or scale S_i as a query. The i-th column shows the feature selection of query scale S_i. The red polygon represents the selection area.

References

1. Caesar, H., Uijlings, J., Ferrari, V.: Coco-stuff: Thing and stuff classes in context. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1209–1218 (2018)
2. 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: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3213–3223 (2016)
3. Mottaghi, R., Chen, X., Liu, X., Cho, N.G., Lee, S.W., Fidler, S., Urtasun, R., Yuille, A.: The role of context for object detection and semantic segmentation in the wild. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 891–898 (2014)
4. Zhou, B., Zhao, H., Puig, X., Xiao, T., Fidler, S., Barriuso, A., Torralba, A.: Semantic understanding of scenes through the ade20k dataset. International Journal of Computer Vision **127**(3), 302–321 (2019)