Supplementary material for HiLo: Exploiting High Low Frequency Relations for Unbiased Panoptic Scene Graph Generation

Zijian Zhou
Department of Informatics
King's College London
zijian.zhou@kcl.ac.uk

Miaojing Shi✉
College of Electronic and Information Engineering
Tongji University
mshi@tongji.edu.cn

Holger Caesar
Intelligent Vehicles Lab
Delft University of Technology
h.caesar@tudelft.nl

To further substantiate the effectiveness of our proposed method, we provide additional experimental results in the supplementary material that could not be included in the paper due to space limitations. The supplementary material includes more experimental results on VG-150.

A. More results on VG-150

We present an extension of our previous results on scene graph detection (SGDet), the most challenging task in the VG-150 [1] dataset, by including comparative evaluations on two simplified tasks: predicate (relation) classification (PredCls) and scene graph classification (SGCls). PredCls involves inferring the relations between objects while assuming prior knowledge of their location and category information. On the other hand, SGCls entails the prediction of the category of an object and the relations between objects, given only their location information. We provide the detailed results of these tasks in the Tab. 1 and Tab. 2.

These results once again demonstrate that our method can improve the performance of low-frequency relations while also taking into account the performance of high-frequency relations. It also shows that the HiLo framework is a general technique that yields systematic improvements in both the panoptic scene graph generation (PSG) and scene graph generation (SGG) tasks.

Method	Predicate Classification	
	R/mR@50	R/mR@100
MOTIF [5]	**64.0** / 15.2	**66.0** / 16.2
+IETrans [6]	54.7 / 30.9	56.7 / 33.6
+HiLo (ours)	53.6 / **33.6**	55.5 / **36.4**
VCTree [4]	**64.5** / 16.3	**66.5** / 17.7
+IETrans [6]	53.0 / 30.3	55.0 / 33.9
+HiLo (ours)	53.4 / **34.0**	55.2 / **37.8**
Transformer [3]	**63.6** / 17.9	**65.7** / 19.6
+IETrans [6]	51.8 / 30.8	53.8 / 34.7
+HiLo (ours)	52.9 / **32.8**	55.9 / **36.1**
GPSNet [2]	**65.1** / 15.0	**66.9** / 16.0
+IETrans [6]	52.3 / 31.0	54.3 / 34.5
+HiLo (ours)	53.3 / **33.8**	55.2 / **37.4**

Table 1. Comparison between our HiLo framework and other methods on PredCls on the VG-150 dataset. Similar to [6], we apply IETrans and our own method on top of four leading baselines.

Method	Scene Graph Classification	
	R/mR@50	R/mR@100
MOTIF [5]	**38.0** / 8.7	**38.9** / 9.3
+IETrans [6]	32.5 / 16.8	33.4 / 17.9
+HiLo (ours)	32.1 / **18.9**	33.1 / **20.9**
VCTree [4]	**39.3** / 8.9	**40.2** / 9.5
+IETrans [6]	32.9 / 16.5	33.8 / 18.1
+HiLo (ours)	35.7 / **21.0**	36.8 / **22.7**
Transformer [3]	**38.1** / 9.9	**39.2** / 10.5
+IETrans [6]	32.6 / 17.4	33.5 / 19.1
+HiLo (ours)	32.3 / **20.1**	33.3 / **22.2**
GPSNet [2]	**36.9** / 8.2	**38.0** / 8.7
+IETrans [6]	31.8 / 17.0	32.7 / 18.3
+HiLo (ours)	31.7 / **18.3**	32.5 / **20.2**

Table 2. Comparison between our HiLo framework and other methods on SGCls on the VG-150 dataset. Similar to [6], we apply IETrans and our own method on top of four leading baselines.

References

[1] Ranjay Krishna, Yuke Zhu, Oliver Groth, Justin Johnson, Kenji Hata, Joshua Kravitz, Stephanie Chen, Yannis Kalantidis, Li-Jia Li, David A Shamma, et al. Visual genome: Connecting language and vision using crowdsourced dense image annotations. *International journal of computer vision*, 123:32–73, 2017. 1

[2] Xin Lin, Changxing Ding, Jinquan Zeng, and Dacheng Tao. Gps-net: Graph property sensing network for scene graph

✉ Corresponding author.

generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3746–3753, 2020. 1

[3] Kaihua Tang, Yulei Niu, Jianqiang Huang, Jiaxin Shi, and Hanwang Zhang. Unbiased scene graph generation from biased training. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3716–3725, 2020. 1

[4] Kaihua Tang, Hanwang Zhang, Baoyuan Wu, Wenhan Luo, and Wei Liu. Learning to compose dynamic tree structures for visual contexts. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6619–6628, 2019. 1

[5] Rowan Zellers, Mark Yatskar, Sam Thomson, and Yejin Choi. Neural motifs: Scene graph parsing with global context. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5831–5840, 2018. 1

[6] Ao Zhang, Yuan Yao, Qianyu Chen, Wei Ji, Zhiyuan Liu, Maosong Sun, and Tat-Seng Chua. Fine-grained scene graph generation with data transfer. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXVII*, pages 409–424. Springer, 2022. 1