

Supplementary Material For Scene Graph Driven Text-Prompt Generation for Image Inpainting

Overview: This supplementary material provides details about the user study and presents additional qualitative results.

A. User Study

To assess the quality of our generation and to compare it with HVITA [1] method, we conducted an experiment on Amazon Mechanical Turk where 172 human workers evaluated the visual quality of object insertion in images. A total of 512 randomly selected images were annotated by 5 different workers. Within each annotation, we send the worker a randomly selected quadruplet: original image, masked object image, generated image by our method and generated image by the baseline method. We asked the worker to tell which among the two generated images is a better representative of original image and which of the two generated objects better fits the scene. Across the 2560 annotations, 68.43% annotators selected our generated image as a better representation of the ground truth and 72.76% selected our generated object as a better fit to the scene.

B. Qualitative Results

Fig 1 and 2 show the generated scene graphs for the different scenes by our method where the red colored node is the predicted object and the connected edges are the predicted relationships. HVITA baseline [1] does not account the relationship between the objects while predicting the masked object whereas our method not only takes the existing relationships into account but also predict the relationships between the masked object and other objects in the scene.

References

- [1] Jiayan Qiu, Yiding Yang, Xinchao Wang, and Dacheng Tao. Hallucinating visual instances in total absentia. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16*, pages 264–282. Springer, 2020. 1, 2, 3

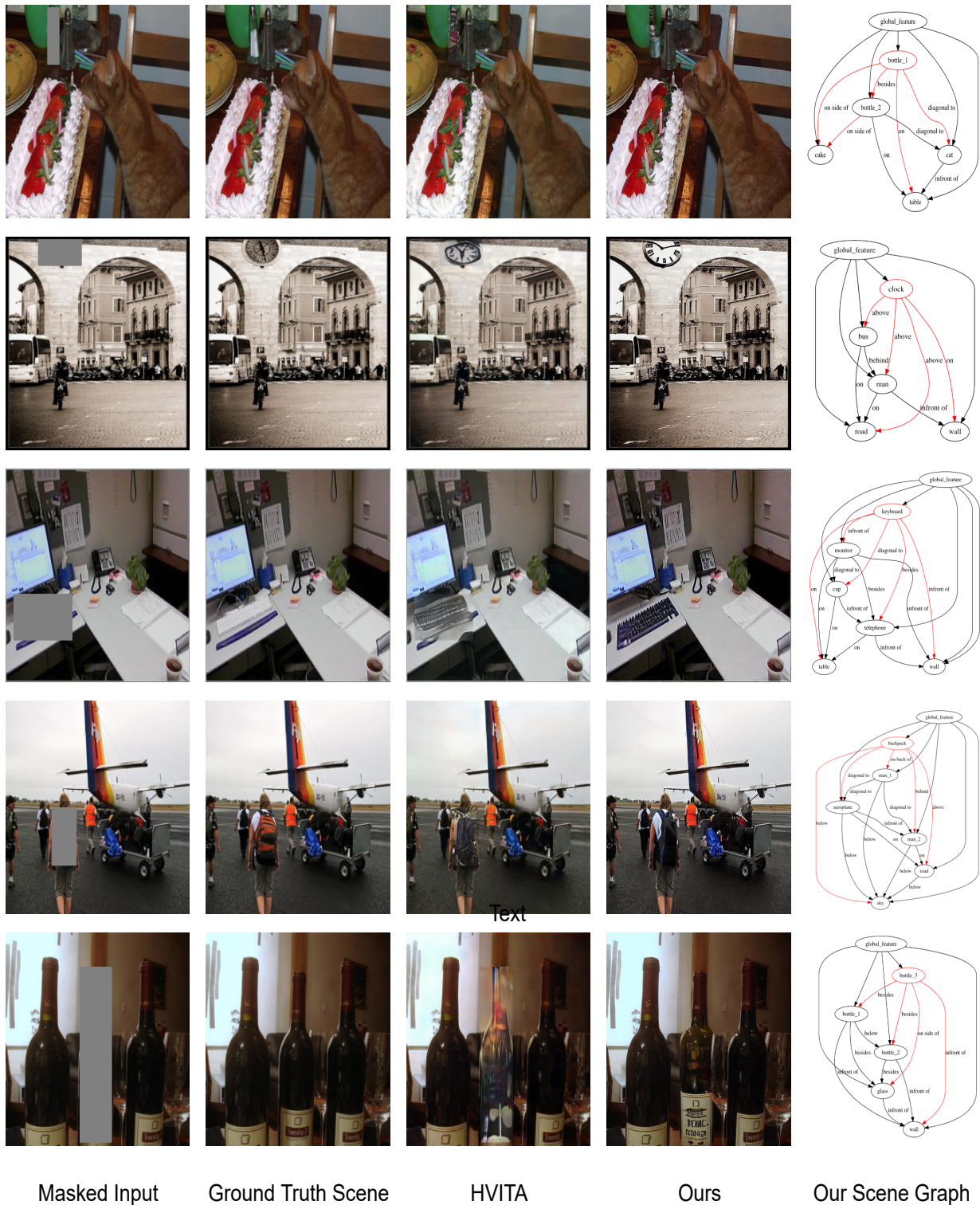


Figure 1. Visual comparison of object generation by our method with HVITA [1] along with generated scene graph by our method for VG-9 dataset.

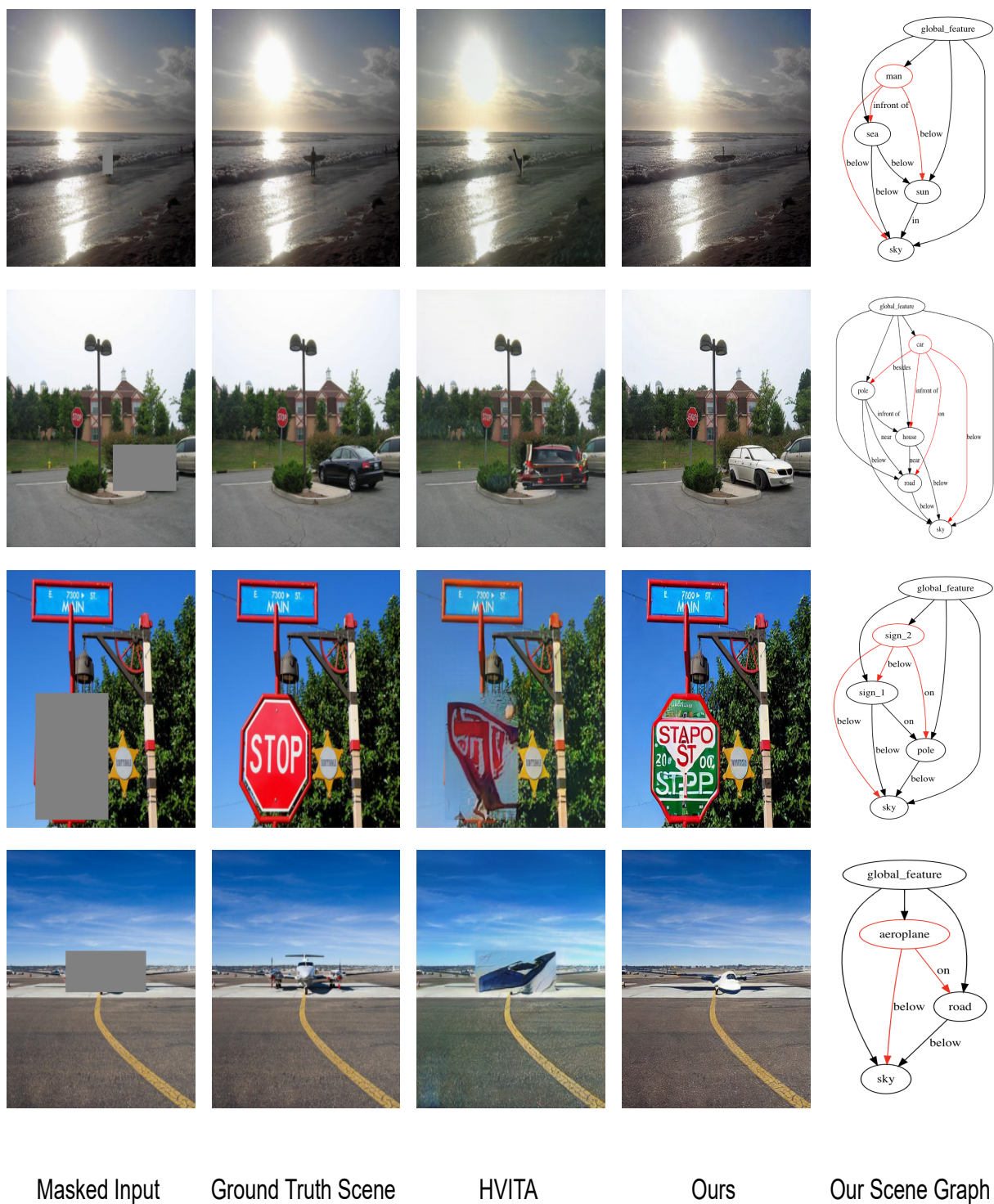


Figure 2. Visual comparison of object generation by our method with HVITA [1] along with generated scene graph by our method for VG-20 dataset.