SUPPLEMENTARY MATERIAL: UNRAVELING VANISHING POINT AND CALIBRATING TINY OBJECTS FOR SEMANTIC SCENE COMPLETION

Anonymous ICIP submission

Author Affiliation

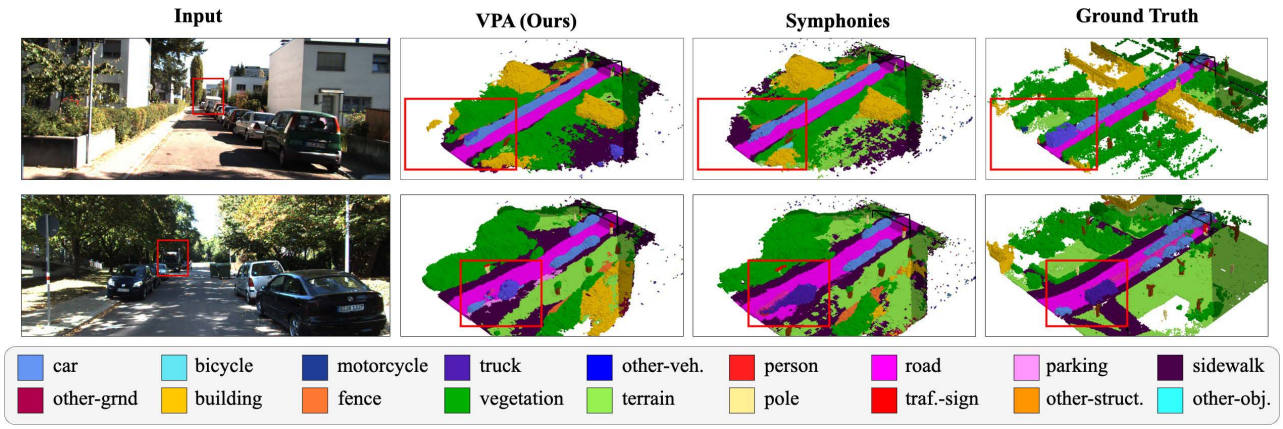

Fig. S1: **Results visualize on SemanticKITTI.** The regions marked with red boxes in the figure demonstrate that our method is capable of accurately detecting and locating small objects, even in challenging scenarios such as long distance predition.

1. OVERVIEW

This document presents additional visualizations and quantitative results on the SemanticKITTI dataset using our Vanish Point Aggregator (VPA) model. In the visualization results section, we provide an example that illustrates how our model enhances the detection and localization of distant and small objects. The visual results include input images, predicted 3D Semantic Scene Completion (SSC) outputs, and the corresponding ground truth, clearly highlighting the improved accuracy in long-distance and small-object perception. In the quantitative results section, our model's performance is evaluated on the validation set of the SemanticKITTI dataset, with comparisons to several SOTA methods. This analysis demonstrates the robustness and effectiveness of our approach in handling challenging scenarios, ultimately improving perception for autonomous driving tasks.

2. VISUALIZATION RESULTS

In this section, we present visualizations that demonstrate the enhanced performance of our Vanish Point Aggregator (VPA) model in detecting small objects and improving long-distance semantic perception. As shown in Fig. S1, the regions marked with red boxes highlight how our method successfully detects and localizes small objects located at long distances, even when they are partially obscured in the 2D camera image. These small objects, shown in the VPA (Ours) column, are segmented with greater accuracy compared to symphonies, which struggles with precise detection at smaller scales. This showcases the capability of our model to effectively handle both long-distance and small-object detection challenges that are often problematic for traditional models.

3. QUANTITATIVE RESULTS

The results in Table S1 show that our method achieves the highest mIoU of 15.26, surpassing all baseline methods. Compared to Symphonies, which attains an mIoU of 14.89, our approach improves by 0.37, demonstrating the effectiveness of integrating VPQ for spatially prioritized feature refinement.

In terms of specific categories, our method achieves the best performance in building 22.01, car 29.03, and vegetation 25.39, indicating its robustness in capturing structured objects and dense areas. Additionally, VPA outperforms others in challenging categories, such as truck 19.27 and bicycle 3.55, where accurate long-range predictions are critical. For smaller objects, such as person 4.07 and traffic sign 5.91, our method achieves competitive performance, benefiting from the fusion of VPQ and instance queries. Although the occupancy prediction slightly decreases compared to some pre-

Method	IoU	mIoU	road	sidewalk	parking	other-grnd.	building	car	truck	bicycle	motorcycle	other-veh.	vegetation	trunk	terrain	person	bicyclist	motorcyclist	fence	pole	traf.-sign
MonoScene [1]	36.86	11.08	56.52	26.72	14.27	0.46	14.09	23.26	6.98	0.61	0.45	1.48	17.89	2.81	29.64	1.86	1.20	0.00	5.84	4.14	2.25
TPVFormer* [2]	35.61	11.36	56.50	25.87	20.60	0.85	13.88	23.81	8.08	0.36	0.05	4.35	16.92	2.26	30.38	0.51	0.89	0.00	5.94	3.14	1.52
VoxFormer-S [3]	44.02	12.35	54.76	26.35	15.50	0.70	17.65	25.79	5.63	0.59	0.51	3.77	24.39	5.08	29.96	1.78	**3.32**	0.00	7.64	7.11	4.18
OccFormer [4]	36.50	13.46	58.85	26.88	19.61	0.31	14.40	25.09	**25.53**	0.81	1.19	8.52	19.63	3.93	32.62	2.78	2.82	0.00	5.61	4.26	2.86
NDC-Scene [5]	37.24	12.70	**59.20**	28.24	**21.42**	**1.67**	14.94	26.26	14.75	1.67	2.37	7.73	19.09	3.51	31.04	3.60	2.74	0.00	6.65	4.53	2.73
H2GFormer-S [6]	**44.57**	13.73	56.08	**29.12**	17.83	0.45	19.74	27.60	10.00	0.50	0.47	7.39	**26.25**	**7.80**	**34.42**	1.54	2.88	0.00	7.24	7.88	4.68
Symphonies [7]	41.92	14.89	56.37	27.58	15.28	0.95	21.64	28.68	20.44	2.54	**2.82**	**13.89**	25.72	6.60	30.87	3.52	2.24	0.00	8.40	**9.57**	5.76
VPA (Ours)	41.99	**15.26**	58.40	27.26	21.21	0.52	**22.01**	**29.03**	19.27	**3.55**	2.55	12.52	25.39	6.20	31.37	**4.07**	2.40	0.00	**9.19**	9.39	**5.91**

Table S1: **Quantitative results on SemanticKITTI `val`.** * represents the reproduced results in [2]. The best results are in **bold**. VPA demonstrates superior performance compared to all other methods, particularly in the score of small objects.

vious methods, our model achieves significantly higher accuracy and precision for small objects. This improvement highlights the advantage of leveraging VPQ to enhance semantic understanding in critical and challenging regions. These results validate the ability of our method to effectively address small object detection and long-range semantic predictions in dense urban scenarios.

4. REFERENCES

[1] Anh-Quan Cao and Raoul De Charette, "Monoscene: Monocular 3d semantic scene completion," in *CVPR*, 2022.

[2] Yuanhui Huang, Wenzhao Zheng, Yunpeng Zhang, Jie Zhou, and Jiwen Lu, "Tri-perspective view for vision-based 3d semantic occupancy prediction," in *CVPR*, 2023.

[3] Yiming Li, Zhiding Yu, Christopher Choy, Chaowei Xiao, Jose M Alvarez, Sanja Fidler, Chen Feng, and Anima Anandkumar, "Voxformer: Sparse voxel transformer for camera-based 3d semantic scene completion," in *CVPR*, 2023.

[4] Yunpeng Zhang, Zheng Zhu, and Dalong Du, "Occformer: Dual-path transformer for vision-based 3d semantic occupancy prediction," in *ICCV*, 2023.

[5] Jiawei Yao, Chuming Li, Keqiang Sun, Yingjie Cai, Hao Li, Wanli Ouyang, and Hongsheng Li, "Ndc-scene: Boost monocular 3d semantic scene completion in normalized device coordinates space," in *ICCV*, 2023.

[6] Yu Wang and Chao Tong, "H2gformer: Horizontal-to-global voxel transformer for 3d semantic scene completion," in *AAAI*, 2024.

[7] Haoyi Jiang, Tianheng Cheng, Naiyu Gao, Haoyang Zhang, Tianwei Lin, Wenyu Liu, and Xinggang Wang, "Symphonize 3d semantic scene completion with contextual instance queries," in *CVPR*, 2024.