Towards Visual Query Localization in the 3D World

Liang Peng[1]*, Bohan Tan[2]*, Zhipeng Zhang[2,3] ‡, Haobo Li[2], Yifan Jiao[4,5], Xingping Dong[1]†, Libo Zhang[4,5]

[1] Wuhan University [2] AutoLab, SAI, Shanghai Jiao Tong University [3] Anyverse Dynamics

[4] University of Chinese Academy of Sciences [5] Institute of Software, Chinese Academy of Sciences

*equal contributions; ‡ first corresponding author; ‡ second corresponding author

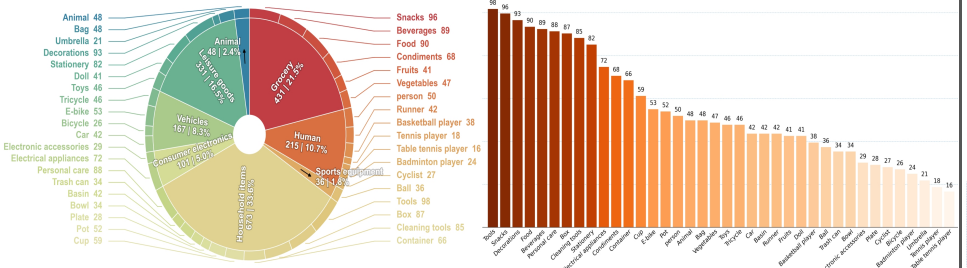
CVPR JUNE 3-7, 2026 DENVER COLORADO

3DVQL

Overview

Contributions

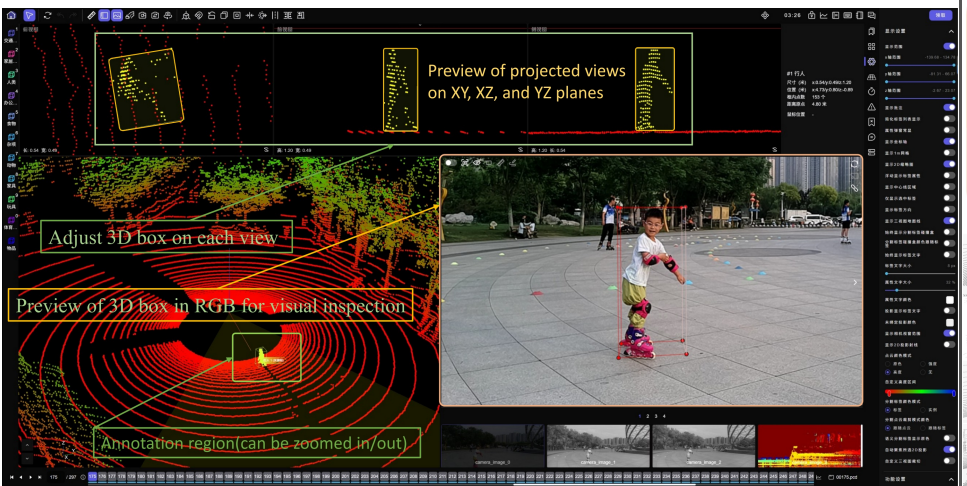
1. A novel 3D visual query localization benchmark, 3DVQL, is introduced, covering RGB, depth, and point cloud data with fine-grained 9-DoF 3D bounding box annotations.
2. A query-guided 3D localization framework is proposed to promote future research on visual query localization in the 3D world.



(a) 8 meta and 38 fine object classes in 3DVQL

(b) The number of sequences for each finely categorized object class

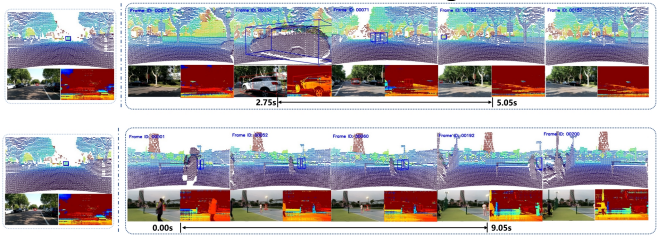
Benchmark	Tot. Seq.	Ann. fr.	Avg trk.	Res trk.	Obj. cls.	Mod. RGB	Mod. PC	Mod. D
2DVQL [10]	2,538	–	90	3.2k	–	✓	✗	✗
3DVQL	**2,002**	**170k**	**84**	**6.4k**	**38**	✓	✓	✓



3DVQL Benchmark

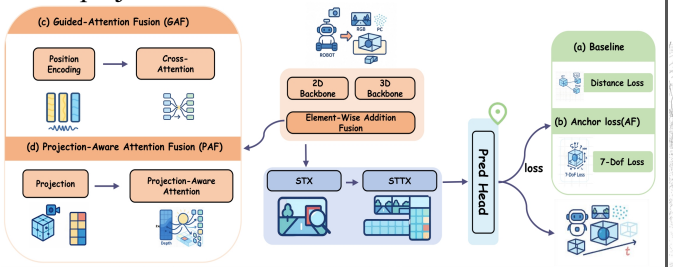
We introduce **3DVQL**, the first benchmark for multimodal 3D visual query localization, with **2,002** sequences, **170K+** frames, **6.4K** response tracks, and **38** object categories.

3DVQL provides aligned point cloud, RGB, and depth modalities, supporting **PC-only**, **RGB-PC**, **RGB-D**, and other multimodal evaluation settings.

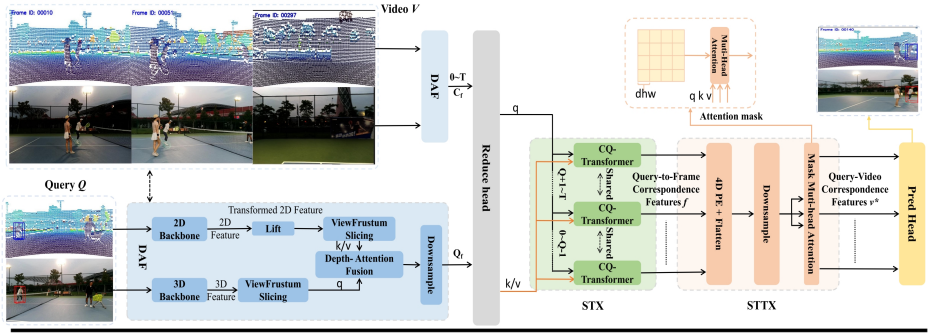


Baselines

We establish three RGB-PC baselines for 3DVQL, including **AF** with anchor-based 3D box prediction, **GAF** with depth-guided attention fusion, and **PAF** with projection-aware **RGB-PC** feature fusion.



LAF: Lift and Attention Fusion



Method	tAP	$tAP_{0.25}$	stAP	$stAP_{0.05}$	rec.%	Succ.
AF	0.181	0.442	0.003	0.015	0.093	11.693
GAF	0.291	0.597	0.015	0.075	0.049	26.309
PAF	0.224	0.577	0.021	0.104	0.115	32.156
LaF	**0.293**	**0.607**	**0.044**	**0.222**	**0.264**	**46.041**

Qualitative Results