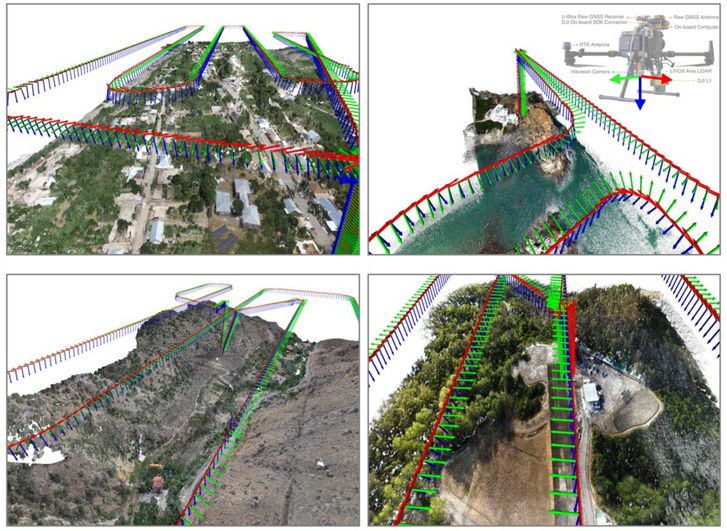
UAVScenes: A Multi-Modal Dataset for UAVs

Sijie Wang[1]* Siqi Li[1]* Yawei Zhang[1]* Shangshu Yu[2]* Shenghai Yuan[1]* Rui She[3]*
Quanjiang Guo[4] JinXuan Zheng[1] Ong Kang Howe[1] Leonrich Chandra[1] Shrivarshann Srijeyan[1] Aditya Sivadas[1] Toshan Aggarwal[1]
Heyuan Liu[1] Hongming Zhang[1] Chujie Chen[1] Junyu Jiang[1] Lihua Xie[1] Wee Peng Tay[1]

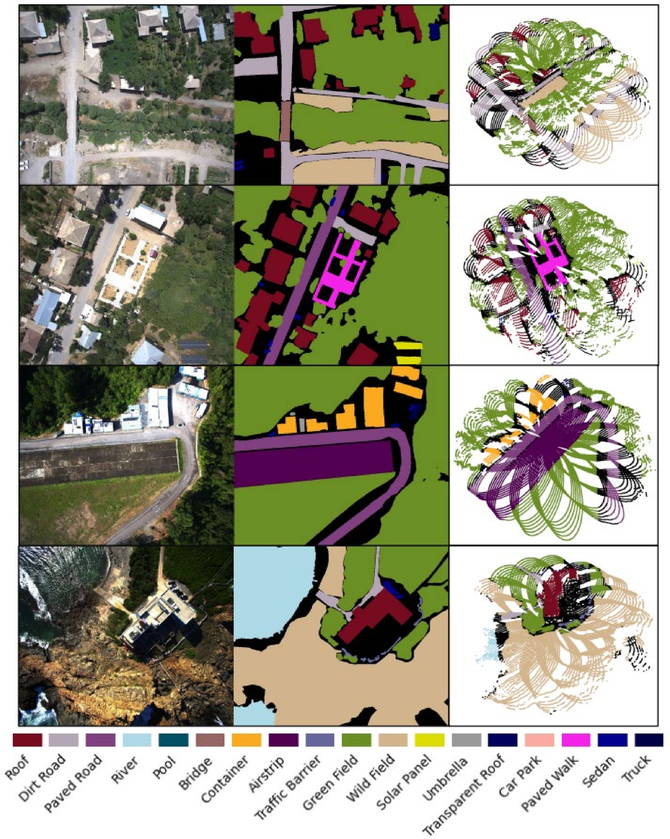
[1]Nanyang Technological University [2]Northeastern University, China [3]Beihang University [4]University of Electronic Science and Technology of China

https://github.com/sijieaaa/UAVScenes

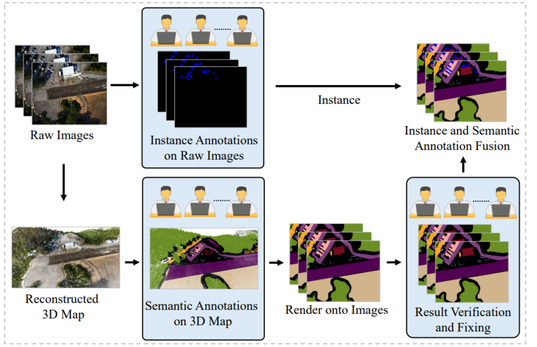
Multi-modal perception is essential for unmanned aerial vehicle (UAV) operations, as it enables a comprehensive understanding of the UAVs' surrounding environment. However, most existing multi-modal UAV datasets are primarily biased toward localization and 3D reconstruction tasks, or only support map-level semantic segmentation due to the lack of frame-wise annotations for both camera images and LiDAR point clouds. This limitation prevents them from being used for high-level scene understanding tasks. To address this gap and advance multi-modal UAV perception, we introduce UAVScenes, a large-scale dataset designed to benchmark various tasks across both 2D and 3D modalities. Our benchmark dataset is built upon the well-calibrated multi-modal UAV dataset MARSLVIG, originally developed only for simultaneous localization and mapping (SLAM). We enhance this dataset by providing manually labeled semantic annotations for both frame-wise images and LiDAR point clouds, along with accurate 6-degree-of-freedom (6-DoF) poses. These additions enable a wide range of UAV perception tasks, including segmentation, depth estimation, 6-DoF localization, place recognition, and novel view synthesis (NVS).



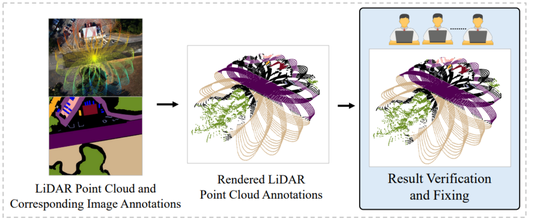
Visualization of the reconstructed 6-DoF poses and 3D mesh/point cloud maps.



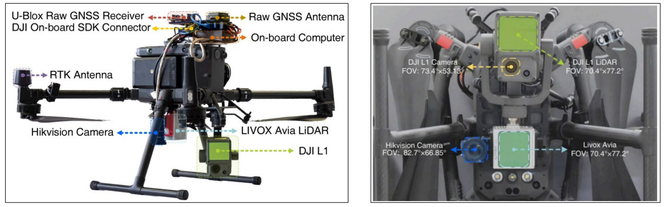
Visualization of frame-wise image and LiDAR point cloud annotations of the proposed UAVScenes dataset.



The 2D image annotating pipeline. Manual annotating is conducted at 3D map annotations, instance annotations, and fixing stages.



The LiDAR point cloud annotating pipeline. Manual annotating is conducted at fixing stages.



The data collection platform used in Mars-LVIG. Images are from Mars-LVIG. The sensors are mounted on the DJI M300 RTK industrial UAV.