CaRF: Enhancing Multi-View Consistency in Referring 3D Gaussian Splatting Segmentation

Tao, Yuwen (School: Shanghai Pinghe Bilingual School)

Referring 3D Gaussian Splatting Segmentation (R3DGS) is a newly proposed task in ICML 2025 Oral paper ReferSplat. Differ from previous 3D segmentation that simply select 3D objects, it emphasized 2D authentic mask generation using texts with adjectives and spatial relationships. ReferSplat itself uses a cross-modal alignment module that shows fantastic performance in terms of spatial reasoning, but after testing, I found that its masks qualities differ a lot in various novel views. To solve this, I introduce Camera-Aware Referring Field (CaRF), a differentiable framework that help models to achieve multi-view consistency during training. To be more specific, this work includes two modules: In-Training Paired-View Supervision (ITPVS) and Gaussian Field Camera Encoding(GFCE). In ITPVS, I utilize two overlapping views to supervise a group of selected Gaussians at a iteration, allowing it to perceive the view differences. In GFCE, I incorporate the encoded camera features in the results of the calculation between text and Gaussians, helping the model the understand the difference introduced in ITPVS using geometric priors. In experiments, my CaRF model surpasses all of the state-of-the-art methods in three well-known datasets Ref-Lerf, Lerf-OVS and 3D-OVS in terms of mIoU(mean IoU) metrics. In conclusion, CaRF improve the 3D understanding of the segmentation models, allowing them to be used in real world scenarios such as embodied AI and AR/VR systems.

Awards Won: