Split&Splat: Zero-Shot Panoptic Segmentation via Explicit Instance Modeling and 3D Gaussian Splatting

Leonardo Monchieri, Elena Camuffo, Francesco Barbato, Simone Milani, Pietro Zanuttigh

University of Padova

Split&Splat is a zero-shot panoptic 3D scene segmentation framework built on 3D Gaussian Splatting (3DGS), designed to bridge the gap between fast, photorealistic scene reconstruction and object-level semantic understanding. While 3D Gaussian splatting excels at high-quality novel-view synthesis, it lacks explicit instance structure and consistent semantics across views. *Split&Splat* addresses these limitations through a two-stage pipeline that explicitly models object instances and embeds full-resolution semantic descriptors without compression.

In the **Split** stage, the method extracts instance masks from multi-view images and enforces cross-view consistency through depth-guided propagation. Depth maps, a sparse COLMAP reconstruction, and an initial dense Gaussian model are jointly used to align 2D segmentations with 3D geometry. This yields refined, view-consistent instance masks and a labeled point cloud, resolving over-segmentation and correcting inconsistencies caused by occlusions and missing detections.

In the **Splat** stage, each object is independently reconstructed via Gaussian Splatting using instance-masked multi-view inputs. The reconstructed objects are then merged into a full scene while refining instance boundaries and suppressing occlusion artifacts. High-dimensional semantic descriptors, extracted from synthetic multi-view renderings, are projected into Gaussians to create a semantically enriched 3D representation. Additional clustering-based semantic correction further separates unintended merges into coherent sub-instances.

Split&Splat produces a panoptic, editable 3DGS scene representation enabling downstream tasks such as instance retrieval, open-vocabulary segmentation, and fine-grained 3D editing. The method achieves state-of-the-art performance on the LERF segmentation benchmark demonstrating both accuracy and efficiency.