

三大创新：

3D一致性实例特征学习、粗到细两级码本离散化、实例级3D-2D CLIP特征关联

2D-3D关联特征不准确：

3DGs采用 α -blending渲染，通过叠加3D点的不透明度权重生成2D像素，这种方式导致：

1.一个2D像素对应多个3D点。

2.一个3D点贡献多个2D像素。

无法建立2D像素与3D点的一一对应关系。

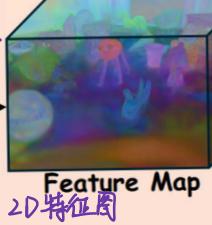
(a) LangSplat
先把3D特征降到2维，再在2D空间做文本匹配

$\mu \ q \ S \ \sigma \ C \ f$



Total Gaussians

rasterize



Rendered image

"old camera"

计算文本查询



cos

Query 2D Mask

Query 2D Mask

依赖深度测试解决遮挡导致多3D点
对称单像素，复杂度高。

完全依赖2D特征图作为中间载体

1. 遮挡部分的3D点无法在2D特征图中体现
2. 2D mask与3D点的关联是间接的(α -blending)



Query image

(b) 直接3D点级查询策略。跳过2D中间层，直接对3D高斯点做文本匹配

$\mu \ q \ S \ \sigma \ C \ f$



Total Gaussians

"old camera"

直接计算文本查询的CLIP特征
与每个3D高斯点的语言特征相似度

Query Gaussians

rasterize



Rendered image (LEGaussians)



Rendered image (LangSplat)



Rendered image (Ours)

Method:

1. 3D一致性实例特征学习 (3D Consistency-Preserving Instance Feature learning)

解决现有方法3D点特征类内一致性差、类间区分度低问题。

基础设置：

1. 给每个3D点新增一个6维实例特征 $f \in \mathbb{R}^6$ ，用于表征实例属性（区别于颜色、位置）

2. 不依赖高维预训练特征 (SAM, CLIP)，也无需跨帧关联的掩码跟踪，仅用 SAM 输出的二值掩码作为监督信号

3. 同一物体高斯点渲染特征相近，不同物体特征远

Loss function:

1. 掩码内平滑损失 (Intra-mask Smoothing Loss)

保证同一物体内部的3D点特征一致

Step:

1) 对任意训练视角，通过3DGs的 α -blending渲染，将所有点的实例特征 f 渲染为2D特征图 $M \in \mathbb{R}^{6 \times H \times W}$

2) 对第*i*个SAM二值掩码 $B_i \in \{0, 1\}^{1 \times H \times W}$ ，计算掩码内所有像素的平均特征 \bar{M}_i

3) 最小化掩码内每个像素特征与平均特征的平方差

$$\mathcal{L}_s = \sum_{i=1}^m \sum_{h=1}^H \sum_{w=1}^W B_{i,h,w} \cdot \|M_{:,h,w} - \bar{M}_i\|^2$$

m - 当前视角 SAM 掩码数量

$B_{i,h,w}$ - 掩码在像素 (h,w) 处取值 (1 或 0)

2. 掩码间对比损失 (Inter-mask Contrastive Loss)

保证不同物体之间的3D点特征差异显著

最大化不同SAM掩码的平均特征之间的距离，避免不同物体的特征混淆

$$\mathcal{L}_c = \frac{1}{m(m-1)} \sum_{i=1}^m \sum_{j=1, j \neq i}^m \frac{1}{\|\bar{M}_i - \bar{M}_j\|^2}$$

$\frac{1}{m(m-1)}$ — 归一化系数

通过倒数，让不同掩码平均特征距离越大，损失越小。

二、两级码本离散化 (Two-level codebook for discretization)

解决连续实例特征的噪声问题，让同一实例的所有3D点拥有完全一致的特征。

1. 粗级码本：

结合实例特征和3D坐标进行聚类，保证空间上相近的3D点归为同一粗聚类。

将每个高斯点的6维实例特征 F 与3维坐标拼接(9维)，初始化大小为 k_1 (64/32) 的粗码本 C_{coarse} 。通过量化将3D点分配到不同粗聚类。

2. 细级码本：

在每一个粗聚类内，仅基于实例特征进一步离散化，提升实例级泛化度。

对每个粗聚类中的3D点，仅使用6维实例特征，初始化大小为 k_2 (5/10) 的细码本 C_{fine} ，最终对每个3D点的特征由粗+细索引确定。

$$\begin{cases} [F \in \mathbb{R}^{n \times 6}; X \in \mathbb{R}^{n \times 3}] \mapsto \{C_{coarse} \in \mathbb{R}^{k_1 \times (6+3)}, I_{coarse} \in \{1, \dots, k_1\}^n\} & \text{coarse, } k_1=64, 32 \\ F \in \mathbb{R}^{n \times 6} \mapsto \{C_{fine} \in \mathbb{R}^{(k_1 \times k_2) \times 6}, I_{fine} \in \{1, \dots, k_2\}^n\} & \text{fine, } k_2=10, 5 \end{cases}$$

3. 伪特征损失：

用1阶段训练的连续实例特征作为伪真值，监督码本离散化过程。

$$L_p = \|M_p - M_c\|'$$

M_p — 第一阶段连续特征渲染的特征图 M_c — 量化特征渲染的特征图 L_p — 损失

三、实例级3D-2D特征关联 (Instance-Level 2D-3D Association without depth test)

建立3D实例与 CLIP 特征的无损关联，避免特征压缩导致语义损失。

对每个3D实例，找到最近配的 SAM 掩码，将该掩码的 CLIP 特征关联到3D实例，无需深度测试，通过特征距离过滤遮挡导致的误匹配。

Step:

(1) 渲染单实例特征图

对每个3D实例（由两级码本索引定义），渲染其“单实例特征图” $M_i \in \mathbb{R}^{6 \times H \times W}$ — 只包含该实例3D点特征，其他区域为0。

(2) 计算 IoU 匹配初始候选

将单实例特征图 M_i 二值化后，与当前视角所有SAM掩码 B_j 计算 IoU，初步筛选出空间重叠度高的 SAM 掩码（候选匹配）。

(3) 特征填充掩码过滤误匹配

针对“遮挡导致多个3D实例与同一SAM掩码高IoU”的问题，引入特征距离验证。

1) 用“-”训练的实例特征填充SAM掩码 B_j ，得到“特征填充掩码” P_j (实例区域为特征均值, else=0)

2) 计算单实例特征图 M_i 与 P_j 的 L_1 距离，越小越好匹配

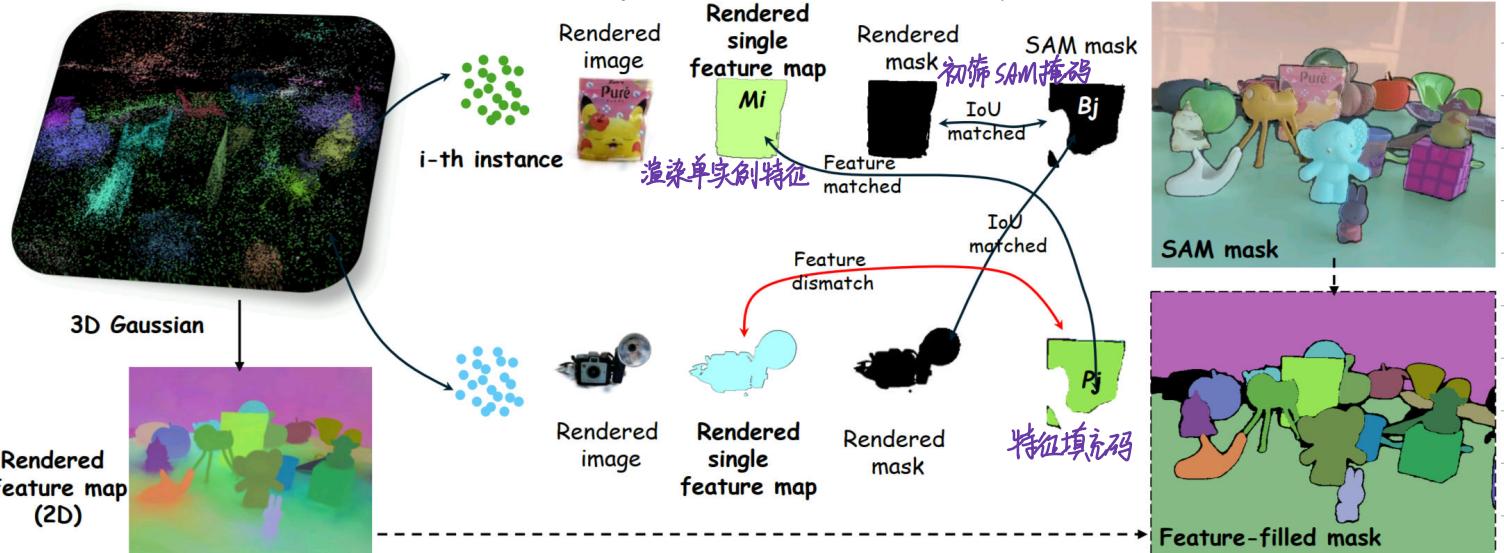
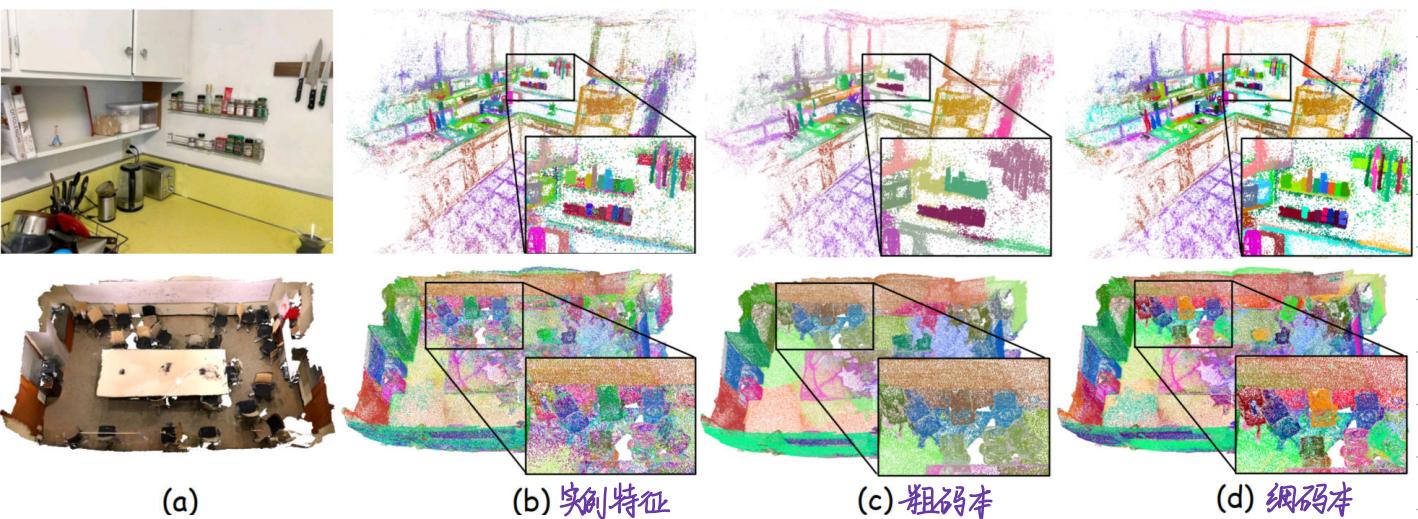
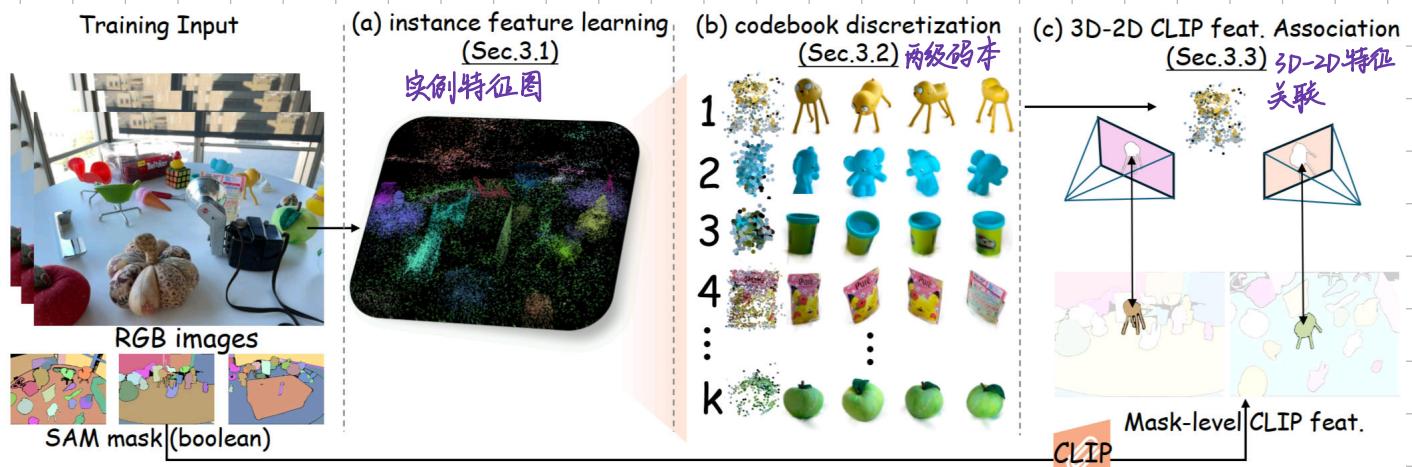
(4) 综合评分与关联 CLIP 特征

定义综合评分 S_{ij} ，结合 IoU (空间匹配) 和 特征距离 (语义匹配)：

$$S_{ij} = \text{IoU}(\pi(M_i), B_j) \cdot (1 - \|M_i - P_j\|_1^1)$$

$\pi(\cdot)$ — 二值化操作，第一项为 IoU (越大越好)，第二项与特征距离成反比

选择评分最高的 SAM 掩码，将其预提取的 CLIP 特征 ($D=512$, 无损) 关联到该3D实例，同时融合多视角的 CLIP 特征。



Conclusion

首个基于3DGFS实现3D点级开放词汇理解的方法

通过SAM掩码训练3D实例特征，设计掩码内、掩码间损失函数

设计粗、细两级码本

提出实例级2D-3D特征关联方法

Limitations :

1. 几何属性与语义内容一致性：高斯点、几何属性不参与实例特征联合优化，导致同一语义类别物体因几何位置分散，未能被精确聚类
 2. 网络超参数 K_1, K_2 对经验有依赖性
 3. 不支持动态