文章编号:1671-4598(2025)09-0302-08 DOI:10.16526/j.cnki.11-4762/tp.2025.09.036 中图分类号:TP391 文献标识码:A

基于三维高斯模型的轻量级分割和编辑方法

江晨炘[1], 禹鑫燚[1,2], 欧林林[1,2]
(1. 浙江工业大学 信息工程学院, 杭州 310023;
2. 浙江工业大学 信息处理与自动化研究所, 杭州 310023)

摘要: 三维场景语义理解作为计算机视觉领域的核心难题之一, 其目标在于实现对三维空间结构的精确识别与分割; 随着无人驾驶、机器人自主导航等应用场景的持续演进, 该任务面临着愈发严苛的挑战标准; 近年来, 3DGS 革新性地被提出, 在保证渲染精度与基线工作相当的前提下, 将重建效率提升数个数量级; 然而当前学术界的探索尚未充分解决基于 3DGS 范式的语义解耦问题; 由于三维数据在复杂度和存储要求等方面都远远超过二维数据集, 高质量标注的三维数据集较为稀缺, 直接训练神经网络理解三维场景语义是困难的; 针对上述挑战, 提出了一种通过二维语义先验知识, 编码低维度信息的办法; 通过预训练的二维语义分割网络提取其中的先验知识, 基于可微体渲染的思想训练一个低维语义信息; 用动态阈值实现语义场粗分割后, 再利用统计学算法滤除噪点重校准; 通过解耦式语义场绑定方案, 实现参数的独立控制; 通过大量实验, 验证了该方法能够通过数秒钟的优化达到之前基线方法的水准, 并能够无缝集成至场景编辑等下游任务。

关键词: 三维高斯模型; 语义分割; 三维场景重建; 扩散模型; 反向传播

Efficient Segmentation and Editing Methods Based on 3D Gaussian Splatting

JIANG Chenxin[1], YU Xinyi[1,2], OU Linlin[1,2]
(1. Engineering of Information, Zhejiang University of Technology, Hangzhou 310023, China;
2. Institute of Information Processing and Automation, Zhejiang University of Technology, Hangzhou 310023, China)

Abstract: 3D scene semantic understanding is a fundamental challenge in computer vision, aiming to accurately recognize and segment three-dimensional spatial structures. With the advancement of applications such as autonomous driving and robotic navigation, this task faces increasingly severe challenges. In recent years, 3D Gaussian splatting (3DGS) is innovatively proposed, which improves reconstruction efficiency by several orders of magnitude while ensuring rendering accuracy and baseline. However, current academic exploration has not fully solved the semantic decoupling based on the 3DGS framework. 3D datasets are more advantages over 3D datasets in high complexity and storage, while high-quality annotated 3D datasets are relatively rare, it is difficult for a directly training network to understand 3D scene semantics. To address this issue, a method is proposed to encode low-dimensional information through 2D semantic prior knowledge, extracting its prior knowledge via a pre-trained 2D segmentation network and training a low-dimensional semantic information with differentiable volume rendering. A dynamic threshold strategy is used to achieve coarse semantic segmentation, and then use statistical algorithms to filter out noise and recalibrate. Furthermore, a decoupled semantic binding approach is introduced for the independent parameter control. Extensive experiments show that this method achieves the baseline performance within a few seconds of optimization and can be seamlessly integrated into downstream tasks such as scene editing.

Keywords: 3D Gaussian model; semantic segmentation; 3D scene reconstruction; diffusion model; back propagation

收稿日期:2025-03-26; **修回日期**:2025-05-06。
基金项目:国家自然科学基金(62373329);浙江省自然科学基金委员会;白马湖实验室联合基金;浙江省自然科学基金重大项目(LBMHD24F030002)。
作者简介:江晨炘(2000-),男,硕士研究生。
欧林林(1980-),女,博士,教授。
通讯作者:禹鑫燚(1979-),男,博士,副教授。
引用格式:江晨炘,禹鑫燚,欧林林. 基于三维高斯模型的轻量级分割和编辑方法[J]. 计算机测量与控制,2025,33(9):302-309.

0　引言

三维场景语义理解是计算机视觉和图形学中极具挑战性且至关重要的任务，这一任务主要涉及从图像或视频中进行场景重建，以及对真实环境的感知两方面。神经辐射场[1]作为一种新兴的场景表达方式，为三维重建领域带来了颠覆性的改变。其核心在于通过深度学习将二维图像序列编码为隐式连续三维场景表示。该技术利用一个神经网络建立空间坐标与视角方向到颜色、体密度的映射关系，结合可微分体渲染实现任意视角的高保真度合成。尽管神经辐射场新颖的架构极大地提升了合成视角的渲染质量，但是同样受限于其完全的隐式表达，这为对其进行语义感知又添加了新的障碍。现有的方法[2-5]往往需要将二维掩码提升到三维，或通过神经场渲染蒸馏语义特征。然而，由于神经辐射场的隐式连续表示，这些方法需要昂贵的随机采样，并且优化计算开销大。此外，神经辐射场的隐式表达这一特性使得在完成语义理解后，依旧难以继续进行下游任务。

三维高斯模型[6]（3DGS，3D gaussian splatting）作为一种新兴的静态场景建模方法逐渐受到关注。3DGS 利用大量彩色的三维高斯通过溅射式栅格化渲染到相机视图中，借助可微渲染和基于梯度的优化对高斯的位置、大小、旋转、颜色和透明度进行精细调节，从而精准还原 3D 场景，具备显式表达能力。对于场景理解而言，基于 3DGS 的分割是一种自然的选择。近期已有工作[7-9]尝试通过将 2D 模型的语义信息升维到 3DGS 中实现分割。但是这些方法依赖于对三维场景的再训练以编码高维语义特征，或需在初始训练阶段并行训练分类网络，这显著增加了显存的要求与计算的开销。针对上述问题，本文提出了一种通过二维语义先验知识，编码低维度信息的办法。通过大量实验，本文验证了该方法能够有效地对三维场景中目标物体进行分割，并证明该方法能够无缝集成至场景编辑等下游任务。

1　三维高斯模型渲染

3DGS 继承了神经辐射场基于体渲染的核心思想，通过半显式建模场景的几何与外观来实现高质量三维重建。与神经辐射场依赖神经网络预测点密度和颜色不同，它将场景表示为数百三维高斯核，每个三维高斯核通过一组参数来刻画局部区域的几何和外观属性，无需再通过耗时的神经场推理与积分计算。渲染时，采用“抛雪球法”[10]，通过将每个高斯基元快速投影到屏幕空间并叠加混合，直接生成像素颜色。这种结合使得高斯泼溅兼顾神经辐射场的高保真细节与实时的渲染效率，同时支持动态场景的高效优化与编辑，成为当前三维重建领域的新标杆。

设 3DGS 中，场景通过如下一个列表表示场景 $\boldsymbol{G}=\{G_0,G_1,\cdots,G_k\}$。对于其中某一三维高斯核 G_i 则可以进一步通过如下一组参数表示：$G_i=\{\mu_i,q_i,s_i,o_i,f_i\}$。其中 $\mu_i\in R^3$ 表示三维高斯核中心的坐标；$q_i\in R^4$ 通过一个四元数存储三维高斯核的朝向；$s_i\in R^3$ 则表示三维高斯核的缩放情况；$o_i\in R$ 表示三维高斯核的透明度；$f_i\in R^{48}$ 通过一个三阶的球谐函数计算在不同观察下，观察三维高斯核时得到的颜色。

在渲染过程中，对于某三维高斯核而言，需要首先通过公式（1）计算其协方差矩阵 $\sum_i$，式中 r_i 为四元数 q_i 对应的旋转矩阵。再通过 u_i 对其进行位移，获得完整的表达式（2）。该表达式相较于常见的三维正态分布而言，缺少了一项归一化系数。作为替代，三维高斯核优化透明度 o_i：

$$\sum\nolimits_i = r_i s_i s_i^T r_i^T \tag{1}$$

$$G_s(x)=\exp\left[-\frac{1}{2}(x_i-u_i)^T\sum\nolimits_i^{-1}(x_i-u_i)\right] \tag{2}$$

对于给定的相机位姿进行光栅化时，需要对三维高斯模型进行视图变换和投影变换，从而与像素空间对齐。视图变换主要是旋转变换和平移变换，它们通常都是仿射的，不会破坏三维高斯核的性质；但投影变换并没有这一性质，为了维持三维高斯核的特性，故通过一个雅可比矩阵 $\boldsymbol{J}$ 对投影变换进行近似。如公式（3）所示，即可得到投影到像素平面上的 $\sum'$：

$$\sum\nolimits_i' = \boldsymbol{J}W\sum\nolimits_i W^T\boldsymbol{J}^T \tag{3}$$

为了减少光栅化的计算成本，3DGS 没有直接对每个像素进行精确计算，而是将整个图像划分成多个不重叠的图块（tile），每个图块包括 16 个像素。接下来，会进一步识别出哪些图块与特定的高斯投影相交。考虑到一个三维高斯核的投影可能会覆盖多个图块，为了加快处理速度，需要复制这些三维高斯核，并为每个新产生的三维高斯核分配一个唯一的标识符，即与之相交的图块的 ID。通过这种方式，3DGS 能有效地降低计算复杂度，同时还保持了图像处理的效率和准确性。

在进行上述预处理之后，即可对每个像素通过公式（4）计算颜色，其中的 c_i 通过 f_i 加权三阶球谐函数获得，α_i 可以通过公式（5）计算：

$$C=\sum_{i\in G}c_i\alpha_i\prod_{j=1}^{i}(1-\alpha_j)=\sum_{i\in G}c_i\alpha_i T_i \tag{4}$$

$$\alpha_i = o_i\times\exp\left[-\frac{1}{2}(x'_i-u'_i)^T\left(\sum\nolimits_i'\right)^{-1}(x'_i-u'_i)\right] \tag{5}$$

在训练阶段，为了优化三维高斯核的参数 $G_i=\{\mu_i,q_i,s_i,o_i,f_i\}$，通过损失函数（6）进行优化，前者

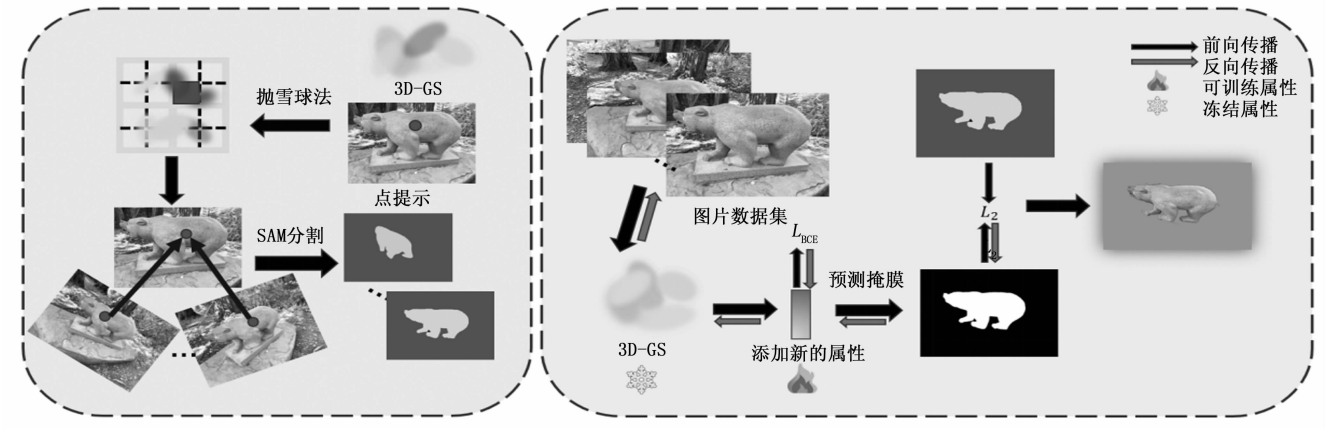

图 1 总体方案流程图

用于保证渲染出来的图片尽可能真实，后者用来提高预测的图片的质量，其中的 λ 为一超参数：

$$L = (1-\lambda)L_1 + \lambda L_{D-SSIM} \tag{6}$$

同时为了更好地表征整个场景，三维高斯模型还采用了一些办法自适应地调节三维高斯核的密度，包括分裂、克隆和剪枝。它们在一定次数的迭代后，针对在视图空间中位置梯度超过特定阈值的高斯执行这些操作，从而提高场景重建的质量。对于克隆，创建高斯核的复制体并朝着位置梯度移动。对于分裂，用两个较小的高斯核替换一个大高斯核，按照特定因子减小它们的尺度。剪枝则是在一定的迭代轮次后，删去透明度 o_i 低于一定阈值的高斯核。

2 轻量化的目标物体分割办法

对于已经预训练完成的三维高斯模型 G 及其对应的训练集 L 中的任意图像，用户可通过在图像上指定提示点 p_0，实现对 3DGS 中目标子集 O 的分割。

由于高质量的三维标注数据较为缺乏，直接识别三维空间结构非常困难。为此，本研究通过为所有三维高斯核添加新的属性 p，基于体渲染的思想，对属性 p 加权求和生成预测掩膜 M_{pred}，并最小化 M_{pred} 与真实掩膜 M_{gt} 之间的误差，分割出一个粗糙的子集 O^*。再通过统计学算法，对粗糙子集 O^* 进行进一步处理，过滤其中的噪点。对于大多数下游任务，现有的扩散模型仍缺乏对整个场景的直接监督能力，本文的方法通过将场景编辑任务解耦为物体编辑任务，显著简化了问题复杂度。

在下游任务中，已经完成与原场景解耦的目标物体可以直接进行一般编辑任务，并不影响原场景的合理性。在外观编辑任务中，本文基于文献［11］的方法，为了更好地诱导扩散模型理解整个场景，本研究引入了 ControlNet[12] 模块，通过引入深度信息，在图像间建立更稳定的控制连接来实现跨视图一致性保障。

2.1 掩膜生成

在直接从单一确定视角出发且缺乏其他有效信息的情况下，直接对三维空间做出语义理解时容易出现歧义。所以需要从单一视角出发，获得多视角下的掩膜。

对于给定的数据集中的视图，首先通过抛雪球法将三维高斯模型 G 投影至该视图中，生成对应的二维高斯核集合 G^*。过滤掉处于像素平面之外和完全不与提示点 p_0 相交的二维高斯核后，保留的集合记为 G_R^*。这些二维高斯的中心位置记为 μ_R^*，则可以通过公式（7）筛选出目标二维高斯：

$$\mu_{sel} = \underset{\mu_R^*}{\operatorname{argmin}}\{|\mu_R^* - p_0|_2 < \xi\} \tag{7}$$

通过公式（7），可以筛选出与提示点距离小于阈值的最近的二维高斯核 μ_{sel}。通过抛雪球法将对应的三维高斯核在整个训练集 L 上投影，生成点提示坐标集合 $p_i(i=1,2,3,\cdots,N)$。接着通过预训练的分割模型，比如 Segment Anything Model[13]（SAM），即可获得掩膜集合，掩膜图像中数值 0 代表背景值，数值 1 代表目标物体。

2.2 目标物体分割

对于预训练完成的三维高斯模型 G，为了防止原先的场景被破坏，首先需要冻结其他所有属性。然后为 3DGS 添加一个新的属性 p_i，用于表征该高斯核属于目标物体的概率。p_i 的训练方式与球谐函数 f_i 类似，但是从不同视角观察某一三维高斯核时，该三维高斯核是否属于目标物体的结果应当是一致的。所以本文将 p_i 设为只有直流分量的球谐函数，从而避免视角的影响。与公式（4）类似，本文通过体渲染的方式计算某一像素上的概率值，进而获得完整的预测掩膜：

$$M_{pred} = \sum_{i\in G} p_i a_i \prod_{j=1}^{i}(1-a_j) \tag{8}$$

为了训练属性 p_i，需要最小化预测掩膜 M_{pred} 与真实掩膜 M_{gt} 之间的误差，本文使用欧氏距离衡量两个变

量之间的差距：

$$L_2 = \sum_{i,j} | M_{\text{pred}}(i,j) - M_{\text{gt}}(i,j) |_2 \tag{9}$$

在物理世界中，一个三维高斯核是否属于目标物体应该为一必然事件或不可能事件，p_i 属性应当接近 0 或者 1。故采用交叉熵函数作为正则项，如公式（10）所示：

$$L_{\text{BCE}} = \sum_{i \in G} BCE(p_i, p_i) \tag{10}$$

最后总的函数如公式（11）所示，其中的 λ 为一超参数：

$$L = L_2 + \lambda L_{\text{BCE}} \tag{11}$$

训练结束后，保留 p_i 大于阈值 γ 的三维高斯核即可初步得到目标物体 O^*。进一步地，本文通过统计学的方法对 O^* 进行过滤筛选得到更加精确的目标物体 O。

Density-Based Spatial Clustering of Applications with Noise[14]（DBSCAN）算法是一种基于密度的经典聚类算法，专为处理含噪声的空间数据结构设计。由于 SAM 模型在分割时，在目标物体边缘处存在一定的噪声，这会影响到本文后续算法的精度，这些三维高斯核往往位于远离目标物体的背景中。为了解决这一问题，本文通过 DBSCAN 算法对点云进行聚类，滤除掉其中的噪点。相较于一般的 DBSCAN 算法而言，本文额外考虑了三维高斯模型中的点云不具有明显的拓扑意义这一特殊性。

对于空间中某点 $\hat{p}$ 而言，其 ε 邻域 $N_\varepsilon(\hat{p})$ 可以定义为：

$$N_\varepsilon(\hat{p}) = \{\overline{q} \in D \mid \text{dist}(\hat{p}, \overline{q}) \leqslant \varepsilon\} \tag{12}$$

称 $\hat{p}$ 为核心点当且仅当其邻域 $N_\varepsilon(\hat{p})$ 满足条件：

$$| N_\varepsilon(\hat{p}) | \geqslant \text{MinPts} \tag{13}$$

其中：$| N_\varepsilon(\hat{p}) |$ 表示集合 $N_\varepsilon(\hat{p})$ 中含有的元素数量，MinPts 为一手动设定的阈值。对于不满足条件（13）的剩余的点，可以根据 $N_\varepsilon(\hat{p})$ 中是否存在至少一个核心点继续分类。$N_\varepsilon(\hat{p})$ 中至少含有一个核心点的点称之为边界点，否则称之为噪点。

对于 DBSCAN 算法而言，其目标是根据给定的超参数 ε 和 MinPts 将整个点云划分为数个簇，从而过滤掉剩下的噪点。考虑到 3DGS 中的离散点云并不具有明确的拓扑结构意义，而且在某些简单区域中，存在少量三维高斯核锚定大量场景的情况。为了减少这些三维高斯核被错误地删除，本文在判定噪声点时，需要再随机挑选 MinPts 个视角，对候选三维高斯核通过相机内外参进行投影，如果在任意视角中，投影都不在预分割掩膜内部，则判定该三维高斯核为噪点，否则记为孤立锚点。对于标记为噪点的三维高斯核，其 p_i 将先被重置为 $\gamma - \gamma^*$。同时，对于一些边界点而言，它们的部分锚定区域，在某些视角的投影中，可以影响到目标物体。对于这部分三维高斯核，如果其 p_i 大于 γ 但是小于 $\gamma + \gamma^*$，则其 p_i 也将暂时被重置为 $\gamma - \gamma^*$。

在使用 DBSCAN 算法筛选出 O^* 的噪点后，由于存在孤立锚点被错误分类的情况，本文继续使用 Delaunay 三角剖分[15]算法进一步搜索空间中被遗漏的三维高斯核。

三角剖分是一个给定点集，生成三角形集合的过程。在二维的情况下，三角剖分是一个通过给定点集，生成三角形集合的过程。对于平面点集 $O^* = \{O_0, O_1, \cdots O_m\}$，希望得到一个三角形集合 $T = \{t_0, t_1, \cdots t_a\}$，该三角形集合满足以下三个条件：1）所有三角形的端点恰好构成点集 O^*；2）任意两个三角形的边不存在相交的情况；3）T 构成集合 O^* 的凸包。满足以上条件的三角形集合 T 往往不唯一，还需要再加上一个质量评定条件加以限制。几种常用的评价指标包括：最小角，即三角形的内角当中角度最小的角；纵横比，即三角形最短边与最长边的比例；半径比，三角形内接圆半径的两倍与外接圆半径的比例。Delaunay 三角剖分算法通常通过空圆准则作为质量评定条件，即对于任意一个三角形，在其外接圆范围内，除了自己本身之外，不包含点集 O^* 中的任何顶点。

在三维的情况下，三角剖分则是一个对于给定的三维点云，生成四面体集合的过程。同上述条件一致，在剖分过程中，四面体需要在满足 3 个条件的同时，通过空球准则作为质量评定条件。

在本文中，考虑到三维高斯点云的特殊性，本文采用 Bowyer-Watson 增量算法实现 Delaunay 三角剖分。首先需要将 DBSCAN 算法中得到的所有边界点纳入到一个足够大的四面体中，然后逐步将这些边界点依次插入到四面体中并重新构造网络。每次插入一个边界点后，对每个新增点实施空球准则验证，删除所有外接球包含该点的四面体单元；在前述操作生成的空腔内，将该点与空腔的三角化表面进行拓扑连接；根据最小最大球准则优化新生四面体形态，确保整体满足空圆准则。在迭代进行上述步骤之后，最终通过优化剔除辅助结构得到一个由 O^* 作为顶点的凸包，将凸包内部阈值大于 $\gamma - \gamma^*$ 的三维高斯核重新加入 O^*。

2.3　场景编辑

经过预训练之后，被分割出的目标 $\gamma - \gamma^*$ 物体能够执行多种下游编辑任务。得益于解耦式场景表示，冻结其余高斯核的属性，仅调整已分割出的三维高斯核，可以实现多样的编辑任务。对于移除目标物体操作而言，可以直接从 3DGS 的列表 $G = \{G_0, G_1, \cdots, G_k\}$ 中移除对

应的三维高斯核，并不会影响场景的其他部分的正确表达。对于目标物体移动或者旋转，可以通过修改 μ_i 中目标物体对应的部分实现。上述两类操作只需要简单微调参数，无需重新训练即可实时生效。

对于外观编辑而言，本文的方案框架如图 2 所示。本文采用语义重构与 3DGS 参数重训练并行的策略。在语义特征编辑层面，通过深度融合 ControlNet[12] 模块的深度引导信息，加强预训练的文生图大模型 Stable Diffusion[16] 的跨模态对齐能力。将深度信息作为条件输入，可以强化图像空间结构表达，从而增强视觉内容与文本提示的语义关联性。在迭代优化过程中，渲染图片在加入噪声送入 Stable Diffusion 进行编辑时，对应的深度图通过 ControlNet 模块一起进行 DDPM 过程，从而增强视觉内容与文本提示的语义关联性。编辑得到的图片逐步更新原始的数据集 3DGS。同时为了保持初始的场景结构不受影响，对应的深度图不做任何更新。该策略既保持了原模型生成质量，又通过深度引导信号实现了对场景几何的细粒度控制，有效提升编辑过程中图像语义与文本指令的匹配精度。

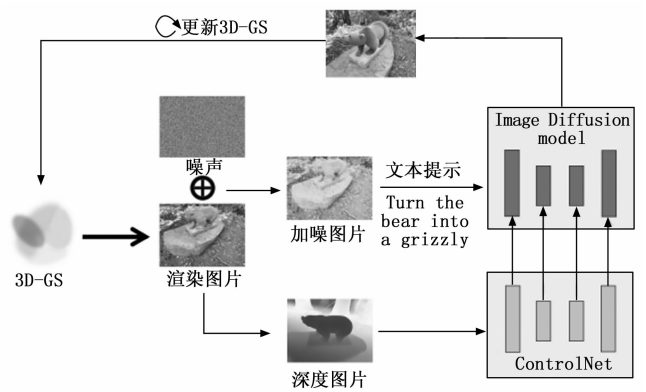

图 2 编辑方案流程图

3 实验结果与分析

3.1 实验基准与实验设置

为验证方法有效性，本文在多个基准数据集上开展评估测试，包括 IN2N、NVOS、Spin-Nerf、Tank&Temples、LERF 和 Mip-NeRF[3, 11, 17-20] 等数据集。这些数据集覆盖室内外环境、前向视角、360 度全景等多种视角的典型场景，充分证明了本方法在各种场景下的泛化能力。

本节使用的定量指标为准确率和交并比。准确率主要衡量真实掩膜与预测掩膜之间的相同部分，准确率越大，可以从一定程度上说明分割的准确程度越高，但是存在一定局限性。交并比 IoU 通过真实掩膜与预测掩膜的交集除以两者之并。同样交并比越大，则说明分割得越准确，同时可以规避一些极端情况，更具有说服力。

在本实验中，实验设置 p_i 学习率为 0.05，使用带有默认参数的 Adam 优化器，超参数 λ_{BCE} 设置为 0.01，分割阈值 γ 为 0.5，γ^* 为 0.05。对于 DBSCAN 算法，其超参数 ε 为 0.1，MinPts 为 20。所有实验均在单张 NVIDIA RTX 3090 GPU 上训练 100 次。对于基线实验，均使用各基线方法论文推荐的默认参数设置和迭代次数，确保了比较的公正性。通过多维度评估实验，本文获取了量化指标与可视化结果的一致性支持，验证了方案的鲁棒性和优越性能。

3.2 定量实验与定性实验结果

本小节首先横向对比了本研究方法和基线方法的定量结果。包括 ISRF、SA3D 以及 SAGA[2, 9, 21] 等基线方法。结果如表 3 所示，本文在两项指标上达到了先进水平。同时本文还对比了 Bear、Garden 和 Family 三个场景中，本文算法与基线算法的平均训练时间和平均显存资源消耗，结果如表 2 所示。综上所述，本文算法可以在资源消耗远少于基线算法的同时，达到相同的分割精度水平。

本文还验证了将给定点提示替换为目标相同的对象的文本提示时本方法的分割能力。对于给定文本输入，通过 DINO[22] 生成目标物体的边界框，将边界框作为 SAM 的输入，以获取二维掩码。然后，通过类似的训练属性 p_i 以及后续处理，可以实现相同的三维目标分割。定量结果见表 1。在测试的场景中，IoU 和 Acc 两项指标均可以达到与使用点提示作为输入提示时同一级别。这些结果表明，本研究的方法可以接受多模态提示作为输入。

表 1 本文算法接受文本输入时的定量结果

场景	Spin/1	Spin/book	Horse	Fortress	Garden
文本	“the box”	“the book”	“the horse”	“the fortress”	“the bonsai”
Acc	97.6	98.1	98.9	97.4	99.2
IoU	91.4	86.3	92.0	93.5	91.9

表 2 本文算法与基线算法的资源消耗比较

	训练时间	最大显存占用
SAGA	18 分钟	15 G
SA3D	7 分钟	10.6 G
Ours	5 秒	4.5 G

图 3（a）展示了本方法二元分割与场景级三维分割的结果。通过在每个场景中对数个分割目标实施双视角渲染的可视化验证，本方法充分展现了基于 SAM 实例掩码预测的三维场景解构能力。实验结果证明，该算法在复杂场景中能精准实现对象级语义解耦与三维重构。

表 3　本文算法与基线算法的定量结果比较

场景	ISRF[21]		SA3D[2]		SAGA[9]		本文算法	
	Acc ↑	IoU ↑	Acc ↑	IoU ↑	Acc ↑	IoU ↑	Acc ↑	IoU ↑
Spin/1	87.9	73.6	97.5	84.8	97.9	83.7	**98.8**	**92.5**
Spin/2	98.8	80.5	**99.7**	92.2	98.8	90.5	99.0	**93.8**
Spin/3	99.0	83.9	**99.8**	**87.4**	93.4	85.9	98.4	85.6
Spin/4	98.9	83.3	**99.8**	**98.3**	99.0	93.3	99.5	96.4
Spin/book	96.3	81.8	**99.7**	**93.1**	98.9	91.8	99.0	84.3
Orchids	90.3	80.4	92.4	84.1	96.3	85.4	**99.7**	**92.1**
Flower	92.2	88.1	88.9	84.6	97.3	88.1	**98.7**	**93.6**
Fortress	91.7	81.4	**98.3**	91.0	96.2	92.0	97.2	**93.7**
Fern	91.4	72.1	95.8	84.7	91.7	86.2	**96.9**	**90.5**
Horns	93.5	73.5	94.6	87.2	94.4	83.1	**98.8**	**92.5**
Garden	96.3	87.5	**99.7**	86.4	93.5	87.1	99.3	**92.1**
Bonsai	96.3	80.4	99.1	**92.3**	96.3	90.4	**99.7**	85.9
Family	82.7	72.6	**99.2**	**93.9**	83.7	80.1	98.8	90.7
Horse	91.3	74.4	99.0	90.5	91.3	84.6	**99.2**	**91.7**
Bear	85.1	77.7	**99.2**	**96.3**	89.2	87.9	97.0	87.2
平均	90.9	79.4	99.0	**90.5**	91.3	84.6	**90.2**	90.1

如图 3（b）所示，展示了本方法的结果在下游任务中的表现，红色边界框中为目标物体。借助 3DGS 的半显式的特性，可以轻易实现对分割后的目标实施各种操作，并保持场景完整性。

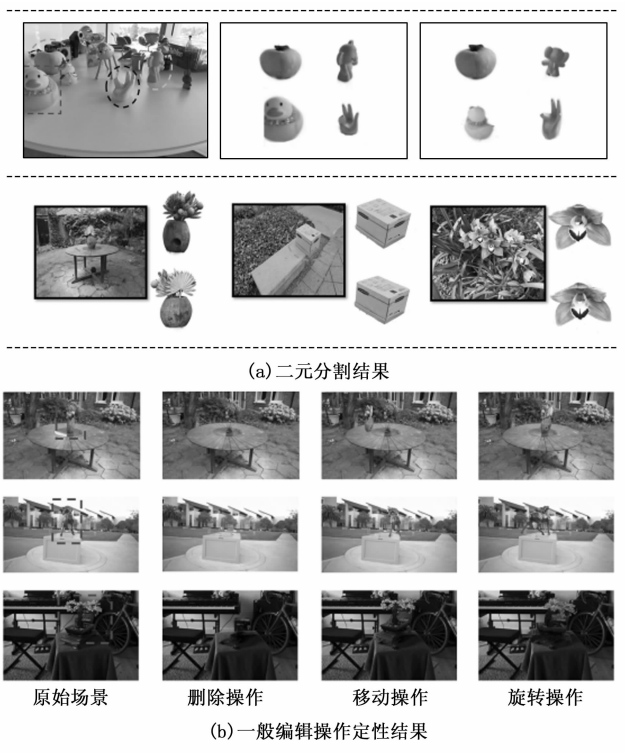

图 3　二元分割结果

如图 4 所示，展现本文在外观编辑方法与基线方法的区别，基线方法虽然在整体上有不错的表现，但是在细节处未能做到完全编辑。相对而言，本文的方法更加稳定和完整。

图 4　外观编辑实验定性结果

3.3　消融实验结果

为验证噪声视角鲁棒性，本实验随机从二维掩码的子集中抽取视角作为对照组，子集中的视图数分别约为原数据集的四分之一、八分之一、十六分之一及三十二分之一。定性验证结果如图 5 所示，实验表明：在两类全景动态场景中，本方法即便仅剩八分之一的视角输入仍可输出完整分割效果；但当训练视角锐减至十六分之一甚至三十二分之一时，算法虽仍能保有一定的分割能力，但不可避免的是，噪声开始增加。

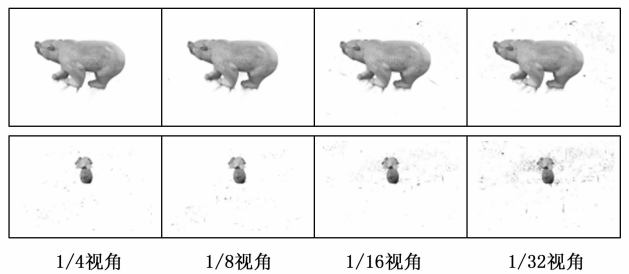

图 5　视角消融实验结果

在本文的方法中，使用交叉熵损失强迫每个三维高斯核属于目标物体的概率靠近 0 或者 1。通过将超参数设置为 0 验证其有效性。实验结果如表 4 所示，验证了包括各种视角在内的 5 个场景，在去除交叉熵正则项之后，所有场景的两项指标平均下降了 1.9% 和 1.5% 左右，这说明交叉熵正则项可以在多数场景中提升本文方法对于目标物体的分割能力。

表 4　交叉熵损失消融实验

场景	使用正则项		不适用正则项	
	Acc	IoU	Acc	IoU
Spin/1	98.8	92.5	96.5	91.1
Spin/2	99.0	93.8	97.0	91.7
Garden	99.3	92.1	96.9	91.4
Horse	99.2	91.7	97.2	90.1
Family	98.8	90.7	98.1	89.4
平均	99.0	92.2	97.1	90.7

随后验证的是 DBSCAN 算法和 Delaunay 算法在本文分割效果中的有效性。DBSCAN 算法的效果如图 6 所示。DBSCAN 算法的引入有效减少了空间中零碎的

噪点，这部分噪点可能是因为 SAM 模型在二维掩膜边界上不够鲁棒导致的。虽然 DBSCAN 算法有效减少了空间中离物体较远的噪点，但是其簇分类效果较依赖于其两个超参数的设置，在 Horse 和 Garden 场景中，目标物体在使用 DBSCAN 算法后均出现了额外的伪影的情况。

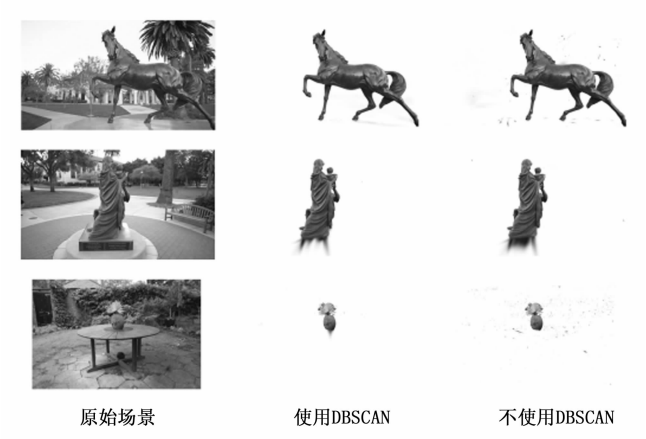

图 6 DBSCAN 消融实验

如图 7 所示，DBSCAN 算法确实出现了将物体内部较为空旷区域处的三维高斯核误删的情况。Delaunay 三角剖分算法很好地将这些点重新归类到目标物体的集合中，减少了编辑后场景中的错误情况。

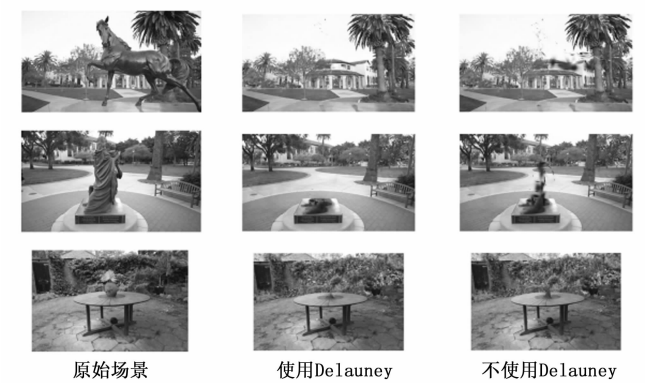

图 7 Delaunay 三角剖分算法消融实验

4 结束语

本文基于体渲染的方法，实现了三维语义分割效能的突破。相较于之前的研究，本文算法不仅在优化效率上获得了提高，同时利用统计学算法，有效过滤被误分类的三维高斯核，提高了本文算法对于复杂背景以及 SAM 模型本身具有的扰动的鲁棒性。技术泛化实验进一步验证了该方案在不同工况下的适配性，在接受文本提示或者视角被减少的情况下，可以返回同样有效且精确的掩码。同时不同场景下的定量实验结果可以证明，本文所提方法可以在下游三维场景编辑任务中展现出良好的性能表现。

参考文献：

[1] MILDENHALL B, SRINIVASAN P P, TANCIK M, et al. Nerf: Representing scenes as neural radiance fields for view synthesis [J]. Communications of the ACM, 2021, 65 (1): 99 - 106.

[2] CEN J, ZHOU Z, FANG J, et al. Segment anything in 3d with nerfs [J]. Advances in Neural Information Processing Systems, 2023, 36: 25971 - 25990.

[3] MIRZAEI A, AUMENTADO-ARMSTRONG T, DERPANIS K G, et al. Spin-nerf: Multiview segmentation and perceptual inpainting with neural radiance fields [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 20669 - 20679.

[4] KIM C M, WU M, KERR J, et al. Garfield: Group anything with radiance fields [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 21530 - 21539.

[5] LIU Y, HU B, Tang C-K, et al. SANeRF-HQ: Segment anything for NeRF in high quality [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 3216 - 3226.

[6] KERBL B, KOPANAS G, LEIMKüHLER T, et al. 3d gaussian splatting for real-time radiance field rendering [C] // Proceedings of Advances in Neural Information Processing Systems, 2023, 42 (4): 14.

[7] YE M, DANELLJAN M, YU F, et al. Gaussian grouping: Segment and edit anything in 3d scenes [C] // Proceedings of Advances in Neural Information Processing Systems, 2023, 42 (4): 14.

[8] QIN M, LI W, ZHOU J, et al. Langsplat: 3d language gaussian splatting [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 20051 - 20060.

[9] CEN J, FANG J, YANG C, et al. Segment any 3d gaussians [C] // Proceedings of the AAAI Conference on Artificial Intelligence, 2025, 39 (2): 1971 - 1979.

[10] ZWICKER M, PFISTER H, VAN BAAR J, et al. EWA volume splatting [C] // Proceedings Visualization, 2001. VIS'01. IEEE, 2001: 29 - 538.

[11] HAQUE A, TANCIK M, EFROS A, et al. Instruct-nerf2nerf: Editing 3d scenes with instructions [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 19740 - 19750.

[12] ZHANG L, RAO A, AGRAWALA M. Adding conditional control to text-to-image diffusion models [C] // Proceedings of the IEEE/CVF International Conference on

Computer Vision，2023：3836－3847.

[13] KIRILLOV A，MINTUN E，RAVI N，et al. Segment anything [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision，2023：4015－4026.

[14] ESTER M，KRIEGEL H-P，SANDER J，et al. Density-based spatial clustering of applications with noise [C] // Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining，Menlo Park：AAAI Press，1996：226－231.

[15] REBAY S J J O C P. Efficient unstructured mesh generation by means of Delaunay triangulation and Bowyer-Watson algorithm [J]. Journal of Computational Physics，1993，106 (1)：125－138.

[16] ROMBACH R，BLATTMANN A，LORENZ D，et al. High-resolution image synthesis with latent diffusion models [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition，2022：10684－10695.

[17] KERR J，KIM C M，GOLDBERG K，et al. Lerf：Language embedded radiance fields [C] // //Proceedings of the IEEE/CVF International Conference on Computer Vision，2023：19729－19739.

[18] BARRON J T，MILDENHALL B，VERBIN D，et al. Mip-nerf 360：Unbounded anti-aliased neural radiance fields [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition，2022：5470－5479.

[19] REN Z，AGARWALA A，RUSSELL B，et al. Neural volumetric object selection [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition，2022：6133－6142.

[20] KNAPITSCH A，PARK J，ZHOU Q-Y，et al. Tanks and temples：Benchmarking large-scale scene reconstruction [J]. ACM Transactions on Graphics (ToG)，2017，36 (4)：1－13.

[21] CHEN X，TANG J，WAN D，et al. Interactive segment anything nerf with feature imitation [J]. Arxiv Preprint Arxiv：2305：16233.

[22] LIU S，ZENG Z，REN T，et al. Grounding dino：Marrying dino with grounded pre-training for open-set object detection [C] //Proceedings of the European Conference on Computer Vision，Cham：Springer Nature Switzerland，2024：38－55.

(上接第 229 页)

[3] WANG X，WANG W，CAI X. A design based on SysML for high lift system of civil aircraft [C] //2021 International Conference on Digital Society and Intelligent Systems (DSInS)，Chengdu，China，2021，pp. 102－106.

[4] RUDOLPH P. High-lift systems on commercial subsonic airliners [R]. 1996.

[5] 关　莉，廉晚祥. 大功率高压直流无刷伺服电机在飞机高升力系统中的应用研究 [J]. 现代制造技术与装备，2020，56 (10)：58－59.

[6] 史佑民，杨新团. 大型飞机高升力系统的发展及关键技术分析 [J]. 航空制造技术，2016 (10)：74－78.

[7] 张新慧，李　晶，任宝平. 大型先进民用飞机高升力控制系统架构研究 [J]. 测控技术，2020，39 (10)：124－129.

[8] KORBACHER G K. Aerodynamics of powered high-lift systems [J]. Annual Review of Fluid Mechanics，1974，6 (1)：319－358.

[9] 王雪鹤，张子瀚，柴春硕，等. 旋翼翼型流动分离特性分析及高升力设计 [J]. 直升机技术，2023 (3)：1－8.

[10] STRIIBER H. The aerodynamic design of the A350 XWB-900 high lift system [C] //29th international congress of the aeronautical sciences，International Council of the Aeronautical Sciences，2014.

[11] 石建强，马高杰. 基于高升力系统襟翼传动线系布局的研究 [J]. 装备制造技术，2023 (8)：63－66.

[12] VAN DAM C P. The aerodynamic design of multi-element high-lift systems for transport airplanes [J]. Progress in Aerospace Sciences，2002，38 (2)：101－144.

[13] QIANF J I，ZHANG Y，HAIXIN C，et al. Aerodynamic optimization of a high-lift system with adaptive dropped hinge flap [J]. Chinese Journal of Aeronautics，2022，35 (11)：191－208.

[14] CHEN G，YANG X，TANG X，et al. Effects of slat track on the flow and acoustic field of high-lift devices [J]. Aerospace Science and Technology，2022，126：107626.

[15] 康　宁，胥海量. 国外宽体客机高升力系统先进作动技术研究 [C] //中国航空学会. 第十届中国航空学会青年科技论坛论文集. 科学普及出版社，2022.

[16] 索晓杰，李亚锋，刘　峰. 正余弦旋转变压器在高升力系统中的应用技术研究 [J]. 航空计算技术，2022，52 (2)：103－106.

[17] 宋建国，刘小周，李子豪. 旋转变压器软解码算法分析研究 [J]. 电子技术应用，2023，49 (5)：62－66.

[18] 马天生，蒙　赟. 高精度旋转变压器设计技术研究 [J]. 微电机，2023，56 (11)：1－6.

[19] 韩建辉，张军红，杜永良. 基于正余弦旋转变压器的襟缝翼角位移传感器信号处理技术研究 [J]. 航空科学技术，2023，34 (3)：82－88.