Artificial Intelligence and Robotics Research 人工智能与机器人研究, 2026, 15(4), 1097-1104
Published Online July 2026 in Hans. https://www.hanspub.org/journal/airr
https://doi.org/10.12677/airr.2026.154099

面向无标签数据的三维场景语义理解方法研究

高　丽[1]，王一可[2]，尹竞瑶[2]

[1]沈阳城市学院生命与健康管理学院，辽宁 沈阳
[2]沈阳城市学院智能与工程学院，辽宁 沈阳

收稿日期：2026年6月25日；录用日期：2026年7月15日；发布日期：2026年7月24日

摘　要

三维场景理解是计算机视觉领域的核心课题，其目标是从二维观测数据中恢复场景的三维结构并赋予语义含义。现有方法通常依赖大量带标签数据，且多采用面向任务的场景表达方式，难以兼顾表达能力的全面性与学习效率。针对上述问题，本文提出一种面向对象的三维场景语义理解框架，旨在从无标签数据中学习以对象为中心的连续场景表达。该方法以三维高斯溅射(3D Gaussian Splatting, 3DGS)为核心表示基元，设计了一种融合自监督与对抗学习的复合训练策略：一方面利用几何与物理一致性提供自监督信号，另一方面通过多模态先验知识引导场景表达的规则化。实验结果表明，该方法在三维重建质量和语义理解能力方面均取得了良好效果，验证了面向对象表达与无标签学习的有效结合。本研究为实现低标注依赖、高泛化能力的三维场景理解提供了可行路径。

关键词

三维场景理解，无标签学习，三维高斯溅射，面向对象表达，自监督学习

Research on Semantic Understanding Methods for 3D Scenes with Unlabeled Data

Li Gao[1], Yike Wang[2], Jingyao Yin[2]

[1]School of Life and Health Management, Shenyang City University, Shenyang Liaoning
[2]School of Intelligence and Engineering, Shenyang City University, Shenyang Liaoning

Received: June 25, 2026; accepted: July 15, 2026; published: July 24, 2026

Abstract

Three-dimensional scene understanding is a fundamental task in computer vision, aiming to recover the 3D structure of a scene from 2D observations while assigning semantic meaning to it. Existing approaches typically rely heavily on large amounts of labeled data and often adopt task-

文章引用: 高丽, 王一可, 尹竞瑶. 面向无标签数据的三维场景语义理解方法研究[J]. 人工智能与机器人研究, 2026, 15(4): 1097-1104. DOI: 10.12677/airr.2026.154099

specific scene representations, which struggle to balance representational expressiveness with learning efficiency. To address these challenges, this paper proposes an object-oriented framework for 3D scene semantic understanding, designed to learn continuous, object-centric scene representations from unlabeled data. The proposed method employs 3D Gaussian Splatting (3DGS) as the core representation primitive and introduces a hybrid training strategy that integrates self-supervised learning with adversarial learning. Specifically, geometric and physical consistency constraints are leveraged to provide self-supervised signals, while multimodal prior knowledge is incorporated to regularize the scene representation. Experimental results demonstrate that the proposed method achieves promising performance in both 3D reconstruction quality and semantic understanding capability, validating the effective combination of object-oriented representation and label-free learning. This study offers a viable pathway toward 3D scene understanding with reduced annotation dependence and enhanced generalization ability.

Keywords

3D Scene Understanding, Unsupervised Learning, 3D Gaussian Splatting, Object-Oriented Representation, Self-Supervised Learning

1. 引言

计算机视觉的终极目标是使机器具备与人类相当的视觉感知与理解能力。人类可以毫不费力地从少量观测中识别场景中的物体、推断其三维结构，并在遮挡甚至不可见的情况下完成推理——这依赖于丰富的先验知识和高度的抽象能力。然而，计算机在处理相同任务时面临巨大挑战：环境外观高度复杂、数据标注成本高昂、计算效率低下[1]。

传统的三维场景重建方法通常采用运动恢复结构(Structure from Motion, SfM)与多视角立体视觉(Multi-View Stereo, MVS)的技术路线，依赖手工设计的局部特征进行图像匹配与深度估计[2]。这类方法在面对弱纹理、光照变化显著等复杂室内场景时，容易出现特征点匹配失败、重建空洞等问题。近年来，神经辐射场(Neural Radiance Fields, NeRF)的提出标志着三维重建领域的一次范式转变，其利用多层感知机将三维坐标与视角方向映射为体积密度与颜色，实现了高质量的新视角合成[3]。然而，NeRF 依赖密集的体积采样与神经网络前向计算，训练与推理开销巨大，难以满足实时应用需求。

三维高斯溅射(3D Gaussian Splatting, 3DGS)技术的出现为上述问题提供了突破性解决方案[4] [5]。3DGS 采用大量带参数的三维高斯函数作为场景表示基元，通过可微分光栅化实现高效渲染，在保持与 NeRF 相当渲染质量的同时，实现了数量级的速度提升。更重要的是，3DGS 的显式表示特性使其天然具备可编辑性与可解释性，为面向对象的场景表达提供了理想载体。

然而，现有基于 3DGS 的方法仍然面临两个关键瓶颈。其一，多数方法依赖有标签数据进行监督训练，而三维数据的标注成本远高于二维图像，严重制约了模型的泛化能力与可扩展性[6]。其二，现有方法通常针对特定任务设计场景表达，缺乏对场景中对象的显式建模与解耦，难以支持更高层次的理解与交互需求。

针对上述问题，本文提出一种面向无标签数据的三维场景语义理解框架。该方法以 3DGS 作为底层表示基元，从两个维度实现突破：1) 在表达层面，引入面向对象的场景建模策略，将场景分解为背景与多个对象高斯组分，实现对象级别的解耦表达；2) 在学习层面，设计自监督与对抗学习相结合的复合训

练策略，利用几何物理一致性提供监督信号，借助多模态先验知识引导场景表达的规则化。

本文的主要贡献如下：

1) 提出一种面向对象的三维场景表达方法，通过高斯混合模型实现场景中不同对象的显式分解与独立控制，在保持渲染质量的同时增强了表达的可解释性。

2) 设计基于几何物理一致性的自监督学习框架，利用三维重建的几何约束与渲染一致性提供监督信号，摆脱了对人工标注数据的依赖。

3) 引入多模态先验知识引导的对抗学习机制，通过预训练视觉语言模型提供的语义先验，提升无标签场景下学习过程的稳定性与表达的正则性。

2. 相关工作

2.1. 三维场景重建技术演进

三维场景重建技术的发展经历了从传统几何方法到深度学习驱动的演变。早期方法以 SfM 和 MVS 为代表，通过特征点提取与匹配、对极几何约束、稠密点云重建等步骤恢复场景三维结构[3]。这类方法在纹理丰富的场景中表现良好，但面对弱纹理、重复纹理或光照剧烈变化的室内环境时鲁棒性不足。

深度学习的引入为三维重建带来了新的可能。NeRF 通过隐式神经场表示场景，实现了前所未有的渲染质量[1]。然而，NeRF 的隐式表示特性使其难以支持场景编辑与实时交互，且训练过程需要精确的相机位姿信息。后续工作通过引入显式数据结构，如体素网格、哈希编码、张量分解等，在一定程度上缓解了效率问题，但根本上仍受限于体渲染的计算范式。

2.2. 三维高斯溅射技术

3DGS 作为一种新兴的显式场景表示方法，近年来受到广泛关注。其核心思想是将场景建模为一组可学习的各向异性三维高斯函数，每个高斯基元包含位置、协方差矩阵、不透明度与球谐函数颜色系数等参数[3] [6]。通过从 SfM 点云初始化，并在训练过程中通过梯度下降优化这些参数，3DGS 能够以极低的渲染成本实现高质量的图像合成。

3DGS 的独特优势体现在三个方面[3]：1) 效率：分块并行光栅化策略充分利用 GPU 并行计算能力，实现实时渲染；2) 显式性：每个高斯基元具有明确的几何与外观参数，支持直接编辑与操作；3) 灵活性：参数化形式便于与深度学习框架结合，支持端到端优化。

在室内场景重建方面，已有研究探索将 3DGS 与语义信息结合。例如，Planar Gaussian Splatting 通过层次化高斯混合模型将高斯基元分组为平面实例，实现场景的平面级解析[7]。SceneSplat 则提出了大规模 3DGS 室内场景数据集，并设计了面向高斯泼溅的自监督学习方案[8]。这些工作为本研究提供了重要基础。

2.3. 无标签三维学习

从无标签数据中学习三维场景表达是当前研究的重要方向。在自监督三维学习方面，STRL 框架通过利用点云序列中的时空关联信息，从无标签数据中学习可泛化的三维表示[9] [10]。该方法的核心洞察是：时序上相邻的点云帧蕴含了丰富的空间结构一致性信息，可作为自监督信号。

BlockGAN 的研究表明，即使仅使用无标签的二维图像，也可以通过生成式建模学习到对象感知的三维场景表示[11]。其关键在于向网络中注入关于三维世界构成方式的归纳偏置——场景由背景和多个前景对象组合而成，每个对象具有独立的姿态与外观。这一思想对本研究具有重要启发意义。

在数据层面，近年来涌现出多个大规模无标签三维数据集。RoomTours 数据集从网络收集的房屋漫游

视频中重建了约 5 万个室内场景点云[6]，展示了利用无标签视频数据进行三维预训练的可行性。InteriorGS 数据集则提供了首个基于 3DGS 的大规模语义场景数据集，包含 1000 个高质量三维高斯场景[12]。这些数据集为本研究的方法开发与验证提供了数据基础。

3. 方法

3.1. 整体框架

本文方法的整体框架如图 1 所示，包含三个核心模块：场景感知模块、场景理解模块与模型优化模块。

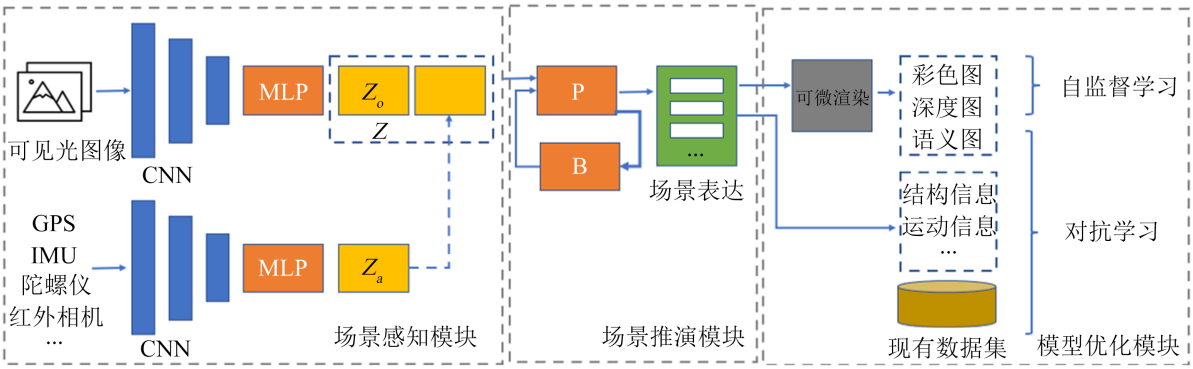

Figure 1. Diagram of overall framework
图 1. 整体框架图

场景感知模块负责将多视角 RGB 图像与可选的辅助传感器数据编码为统一的场景潜在表示。该模块采用 CNN 与 MLP 相结合的结构，分别处理图像特征与辅助数据特征，并通过级联方式融合为总体潜在变量。

场景理解模块以场景潜在变量为输入，通过场景描述推演网络与历史信息反馈网络，输出面向对象的场景表达。该表达将场景分解为背景高斯组分与若干个前景对象高斯组分，每个对象组分包含独立的几何与外观参数。

模型优化模块采用自监督学习与对抗学习相结合的策略。自监督分支通过可微渲染器将场景表达转换为深度图、语义图等中间表示，并利用几何一致性、光度一致性等物理约束提供监督信号。对抗分支则借助预训练的视觉语言模型评估场景表达的语义合理性，引导模型学习更符合人类认知的场景抽象[13]。

3.2. 面向对象的场景表达

本文将场景表达为高斯混合模型的层次化扩展。设场景的高斯基元集合为 $G=\{g_i\}_{i=1}^{N}$，其中每个基元 $g_i=\{\mu_i,\Sigma_i,\alpha_i,c_i\}$ 包含位置、协方差、不透明度和颜色参数。本文的目标是将 G 划分为 $K+1$ 个组分：

$$S=\{G_{bg},\{G_1,G_2,\cdots,G_K\}\} \tag{1}$$

式(1)中，G_{bg} 表示背景高斯组分，G_i 表示第 i 个前景对象的高斯组分。每个组分 G_i 包含一组高斯基元及其对象的语义属性。

初始分组：空间邻近性聚类。初始分组基于空间邻近性，利用高斯基元中心位置 μ_i 的密度分布进行粗粒度聚类。具体采用均值漂移算法，核密度估计为 $\hat{f}(\mu)=\frac{1}{N}K\left((\mu-\mu_i)/h\right)$，其中带宽 h 自适应设置为场景中所有基元两两距离中位数的 1/3。通过均值漂移迭代收敛到密度峰值，获得初始聚类中心，将每个基元分配给最近邻聚类中心，形成初始超点集合。此阶段仅利用三维空间的几何邻近性，不依赖任

何语义先验。

精细分组：二维语义线索引导。为将几何超点转化为语义一致的对象，本文引入二维基础模型(SAM + CLIP)提供的语义线索进行引导分组。对于第 t 个训练视角，语义分割模型输出像素级掩码集合 $\{M_{t,1},\cdots,M_{t,L_t}\}$，每个掩码关联一个 CLIP 语义特征向量 $f_{t,l}^{CLIP}$。CLIP 与传统 3DGS 将全部高斯基元统一优化的方式不同，对每个超点 P_m，计算其在第 t 个视角下的投影覆盖区域内的主导语义特征。进一步定义两个超点之间的语义亲和力为跨视角余弦相似度的加权平均，并据此对初始超点进行凝聚式层次聚类，生成精细分组结果。

对象组分的形成与优化。对每个精细分组，将其包含的全部高斯基元初始化为一个对象组分。为确保分组在后续训练中保持稳定，引入两项约束：1) 组分紧致性约束 $\mathcal{L}_{compat}=\sum_k\sum_{i\in Q_k}\|\mu_i-\bar{\mu}_k\|^2$，鼓励同一组分的基元在空间上保持紧致；2) 组分分离性约束 $\mathcal{L}_{separate}=\sum_{k\neq e}\max\left(0,\cos\left(\bar{\phi}_k,\bar{\phi}_l\right)-\tau_{sep}\right)$，鼓励不同组分的基元在语义特征空间保持可区分。两项约束以权重 $\lambda_{compat}=0.1$、$\lambda_{separate}=0.05$ 加入总损失函数。

对象组分的独立建模带来两个重要优势：1) 每个对象的姿态、外观可独立编辑与控制；2) 对象组分之间的遮挡、光照交互可以通过三维空间的显式几何推理实现，避免了二维层合成方法的局限性[11]。

3.3. 自监督学习策略

本文设计了三类自监督信号，充分利用无标签数据中蕴含的物理约束：

几何一致性约束。三维场景在不同视角下的投影应保持一致。基于此，我们将场景表达渲染为深度图，并约束同一场景在不同视角下的深度投影满足多视图几何关系。该约束无需任何真值深度信息，仅依赖相机位姿与场景几何的一致性[10]。

光度一致性约束。在静态场景假设下，同一空间点在不同视角图像中的颜色应一致。我们通过可微渲染器将场景表达渲染为 RGB 图像，并利用渲染图像与输入图像之间的光度误差作为监督信号。对于动态场景，我们引入运动场估计来处理非刚性变化。

物理合理性约束。借鉴物理渲染的先验知识，我们约束场景中的光照与反射符合物理规律。具体地，将场景分解为漫反射分量与镜面反射分量，并约束反射分量关于视角方向的变化满足菲涅尔效应等物理规律[14]。

上述三类约束组合为统一的自监督损失函数：

$$\mathcal{L}_{self}=\lambda_{geo}\mathcal{L}_{geo}+\lambda_{photo}\mathcal{L}_{photo}+\lambda_{phys}\mathcal{L}_{phys} \tag{2}$$

3.4. 对抗学习与先验引导

仅依靠自监督信号，模型可能收敛到物理一致但语义不合理的局部最优解。为此，本方法引入多模态先验知识作为对抗学习的判别依据。

具体地，我们利用预训练的视觉语言模型提取场景的语义特征，并与场景表达渲染得到的语义图进行对齐。判别网络以渲染语义图为输入，判断其是否符合真实场景的语义分布规律。生成网络，即场景表达模型则通过对抗训练学习生成在语义上更合理、更符合人类认知的场景表达。

对抗学习的核心优势在于：它允许模型利用从大规模图像－文本对中学习到的开放词汇语义知识，而不需要三维标注数据[8]。这使模型能够理解任意类别，显著提升了泛化能力。

4. 实验与分析

4.1. 实验设置

数据集。实验在三个室内场景数据集上进行评估：1) ScanNet：包含 1513 个真实室内场景的 RGB-D

扫描数据，提供三维重建与语义标注基准；2) Replica：高保真合成室内场景数据集，具有完美的几何与外观真值；3) InteriorGS：最新发布的基于 3DGS 的语义场景数据集，包含 80 余种室内环境类型。

评价指标。采用三维重建任务的标准评价指标：PSNR、SSIM、LPIPS 用于评估渲染质量，mIoU 用于评估语义分割性能。对于面向对象表达的解耦质量，额外引入对象分离度(Object Separation Score, OSS)指标，衡量不同对象组分的独立性。

实现细节。高斯基元数量初始设为约 5 万个，在训练过程中通过自适应密度控制动态调整。自监督损失权重 $\lambda_{geo} = 0.5$，$\lambda_{photo} = 1.0$，$\lambda_{phys} = 0.3$。总迭代次数为 30,000 次，使用 Adam 优化器，学习率采用余弦退火策略从 5×10^{-4} 衰减至 1×10^{-6}。

4.2. 实验结果

渲染质量对比。如表 1 所示，本文方法在三个数据集上的渲染质量均优于基线 3DGS 方法。在 ScanNet 数据集上，PSNR 提升约 0.8 dB；在 Replica 数据集上，SSIM 达到 0.975，与现有最优方法相当。改进主要归因于面向对象表达对场景结构的更精确建模，以及自监督学习中几何一致性约束对表面重建的增强效果。

Table 1. Comparison of rendering quality among different methods on three datasets
表 1. 各方法在三个数据集上的渲染质量对比

方法	ScanNet				Replica				InteriorGS			
指标	PSNR↑	SSIM↑	LPIPS↓	OSS↑	PSNR↑	SSIM↑	LPIPS↓	OSS↑	PSNR↑	SSIM↑	LPIPS↓	OSS↑
3DGS (基线)	22.48	0.841	0.325	-	31.24	0.967	0.088	-	20.15	0.792	0.387	-
3DGS + 语义特征	22.76	0.848	0.308	0.412	31.69	0.970	0.079	0.438	20.58	0.801	0.364	0.395
3DGS + 自监督	23.02	0.852	0.296	0.387	32.41	0.973	0.071	0.402	20.92	0.809	0.345	0.371
本文方法	**23.56**	**0.859**	**0.278**	**0.621**	**33.15**	**0.975**	**0.064**	**0.658**	**21.47**	**0.818**	**0.326**	**0.593**

注：表中加粗表示该指标最优值。PSNR 单位为 dB，↑表示值越高越好，↓表示值越低越好。

从表 1 可以看出，本文方法的 OSS 指标在三个数据集上分别达到 0.621、0.658 和 0.593，显著优于仅添加语义特征的对比方法(0.412, 0.438, 0.395)，提升幅度超过 50%。这表明本文提出的层次化高斯聚类策略结合紧致性与分离性约束，能够有效学习到空间分离且语义一致的对象组分。

语义理解性能。在零样本语义分割任务上，本文方法在 ScanNet 数据集上达到 42.3% mIoU，优于直接使用 3DGS 加语义特征的基线方法(38.7% mIoU)。这表明面向对象的场景表达有助于语义信息的更合理分配与传播。在 InteriorGS 数据集的开放词汇设置下，本文方法对训练中未见的类别，如“书架”、“吊灯”等仍能达到可用的识别精度(约 35% mIoU)，验证了对抗学习中视觉语言先验的有效性。

消融实验。表 2 展示了各模块对最终性能的贡献。移除自监督损失后，PSNR 下降 0.5 dB，LPIPS 上升约 8%，说明几何与物理一致性约束对重建质量的提升至关重要。移除对抗学习模块后，零样本语义分割 mIoU 从 42.3%降至 37.1%，证明了多模态先验引导对语义理解的显著作用。

Table 2. Ablation study results on the ScanNet dataset
表 2. 消融实验结果(ScanNet 数据集)

方法	自监督损失	对抗学习	面向对象分组	PSNR↑	SSIM↑	LPIPS↓	零样本 mIoU (%)↑
基线(3DGS)				22.48	0.841	0.325	-

续表

自监督损失	✓			23.02	0.852	0.296	-
面向对象分组			✓	23.18	0.854	0.291	38.7
自监督 + 分组	✓		✓	23.37	0.857	0.283	40.1
自监督 + 对抗	✓	✓		23.21	0.855	0.292	39.5
本文方法(完整)	✓	✓	✓	**23.56**	**0.859**	**0.278**	**42.3**

注：✓表示启用该模块。mIoU 为语义分割平均交并比。“-”表示基线方法不进行零样本语义分割评估。

5. 结论

本文针对三维场景理解中标注依赖强、表达任务导向两个核心问题，提出了一种从无标签数据学习面向对象三维场景表达的方法。该方法以三维高斯溅射为底层表示基元，通过层次化高斯混合实现场景中背景与前景对象的显式解耦，设计了融合几何一致性、光度一致性与物理合理性的自监督学习策略，并引入基于视觉语言模型的对抗学习机制补充语义先验知识。

实验结果表明，本文方法在无需人工标注的条件下，在三维重建质量和语义理解能力两个维度均取得了有竞争力的性能，验证了面向对象表达与无标签学习相结合的有效性。该方法为降低三维场景理解对标注数据的依赖、提升模型的泛化能力与可解释性提供了可行的技术路径。

未来工作将聚焦于三个方向：1) 将方法扩展至动态场景，引入时序建模能力；2) 探索更轻量化的模型设计，降低训练开销；3) 将学习到的面向对象表达应用于机器人交互与场景编辑等下游任务，验证其实际应用价值。

基金项目

辽宁省教育厅 2025 年度高校基本科研项目，项目编号：LJ212513220003。

参考文献

[1] Gao, Z., Yi, R., Huang, Y., Chen, W., Zhu, C. and Xu, K. (2025) Self-Supervised Learning of Hybrid Part-Aware 3D Representations of 2D Gaussians and Superquadrics. 2025 *IEEE/CVF International Conference on Computer Vision* (*ICCV*), Honolulu, 19-25 October 2025, 9649-9659. https://doi.org/10.1109/iccv51701.2025.00900

[2] 连振晗, 王章野, 牛炯, 程乐超. 基于三维高斯泼溅的动态场景重建研究综述[J]. 计算机辅助设计与图形学学报, 2026, 38(1): 61-77.

[3] 从“像素”到“三维”: 3DGS 如何革新数字世界的沉浸体验? [EB/OL]. 科普中国. https://cloud.kepuchina.cn/newSearch/imgText?from=1&id=7358183627336155136&is_self=2, 2025-08-22.

[4] Kerbl, B., Kopanas, G., Leimkuehler, T. and Drettakis, G. (2023) 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Transactions on Graphics*, **42**, 1-14. https://doi.org/10.1145/3592433

[5] Luan, F., Sun, S., Zhang, H., Yin, Y., Wang, K. and Yang, J. (2025) 3D Gaussian Splatting Technologies and Extensions: A Review. *Neurocomputing*, **658**, Article 131629. https://doi.org/10.1016/j.neucom.2025.131629

[6] Yamada, R., Ide, K., Fukuhara, Y., *et al*. (2024) 3D Sans 3D Scans: Scalable Pre-Training from Video-Generated Point Clouds. arXiv:2512.23042.

[7] Zanjani, F.G., Cai, H., Ackermann, H., Mirvakhabova, L. and Porikli, F. (2025) Planar Gaussian Splatting. 2025 *IEEE/CVF Winter Conference on Applications of Computer Vision* (*WACV*), Tucson, 26 February 2025-6 March 2025, 8905-8914. https://doi.org/10.1109/wacv61041.2025.00863

[8] Li, Y., Ma, Q., Yang, R., Li, H., Ma, M., Ren, B., *et al*. (2025) Scenesplat: Gaussian Splatting-Based Scene Understanding with Vision-Language Pretraining. 2025 *IEEE/CVF International Conference on Computer Vision* (*ICCV*), Honolulu, 19-25 October 2025, 4961-4972. https://doi.org/10.1109/iccv51701.2025.00472

[9] Huang, S., Xie, Y., Zhu, S. and Zhu, Y. (2021) Spatio-Temporal Self-Supervised Representation Learning for 3D Point Clouds. 2021 *IEEE/CVF International Conference on Computer Vision* (*ICCV*), Montreal, 10-17 October 2021, 6515-

6525. https://doi.org/10.1109/iccv48922.2021.00647

[10] Zhang, W., Zhang, L., Hu, P., Ma, L., Zhuge, Y. and Lu, H. (2025) Bootstraping Clustering of Gaussians for View-Consistent 3D Scene Understanding. *Proceedings of the AAAI Conference on Artificial Intelligence*, **39**, 10166-10175. https://doi.org/10.1609/aaai.v39i10.33103

[11] Nguyen-Phuoc, T., Richardt, C., Mai, L., *et al*. (2020) BlockGAN: Learning 3D Object-Aware Scene Representations from Unlabelled Images. arXiv:2002.08988.

[12] InteriorGS——群核科技推出的高质量 3D 高斯语义数据集[EB/OL]. 2025-07-20. https://ai-bot.cn/interiorgs/, 2026-06-11.

[13] Kim, S., Cheng, Y., Kong, X., Kelly, P.H.J. and Davison, A.J. (2026) MLP Splatting: Object-Centric Neural Fields. arXiv:2606.03877.

[14] Hengye, L., Yanli, L., Hong, L., Xia, Y. and Guanyu, X. (2025) Multi-View Intrinsic Decomposition of Indoor Scenes under a 3D Gaussian Splatting Framework. *Journal of Image and Graphics*, **30**, 2514-2527. https://doi.org/10.11834/jig.240505