

# 基于多通道Mamba自适应高斯溅射网络的高精度单视图3D重建

周三琪

上海理工大学出版学院, 上海

收稿日期: 2025年8月17日; 录用日期: 2025年9月9日; 发布日期: 2025年9月17日

## 摘要

因缺乏多视角信息, 现有单视图三维重建方法普遍难以兼顾建模精度与泛化能力。为此, 本文提出一种多通道Mamba自适应高斯溅射网络(MMAGS)来实现高精度重建。该方法在网络结构上引入多通道并行Vision Mamba模块, 通过通道划分与并行驱动提升遮挡区域的几何还原能力与跨视角结构信息的稳健提取。为增强纹理细节与图像边缘保真度, 进一步提出深度-颜色双梯度感知的自适应三维高斯滤波策略, 依据局部几何与纹理变化动态调整高斯协方差。所提出的方法应用于ShapeNet-SRN与CO3D两类数据集。实验结果表明, 所提出的MMAGS方法展现出比现有最先进方法更高的重建精度和重建效果, 具备良好的跨类别泛化能力与真实场景鲁棒性。

## 关键词

单视图三维重建, 3DGS, Vision Mamba, 自适应滤波

# Multi-Channel Mamba Adaptive Gaussian Splatting Network for High-Precision Single-View 3D Reconstruction

Sanqi Zhou

College of Publishing, University of Shanghai for Science and Technology, Shanghai

Received: Aug. 17<sup>th</sup>, 2025; accepted: Sep. 9<sup>th</sup>, 2025; published: Sep. 17<sup>th</sup>, 2025

## Abstract

Single-view 3D reconstruction faces significant challenges due to the lack of multi-view constraints,

often leading to a compromise between modeling accuracy and cross-category generalization. We propose MMAGS, a Multi-channel Mamba Adaptive Gaussian Splatting Network designed for high-precision single-view 3D reconstruction. This framework incorporates a multi-branch Vision-Mamba backbone that segments intermediate features into parallel channels, facilitating robust structural reasoning and improved geometric recovery in occluded and texture-limited areas. To enhance detail preservation further, we introduce a depth-color dual-gradient adaptive 3D Gaussian filter that dynamically adjusts Gaussian covariance according to local geometric and photometric variations, thereby significantly enhancing edge sharpness and texture fidelity. We evaluate MMAGS on the ShapeNet-SRN and CO3D benchmarks, where it demonstrates state-of-the-art reconstruction accuracy and perceptual quality, exhibiting strong cross-category generalization and robustness across both synthetic and real-world datasets.

## Keywords

Single-View 3D Reconstruction, 3D Gaussian Splatting, Vision Mamba, Adaptive Gaussian Filtering

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

单视图三维重建旨在从单张二维图像推断出场景或物体的三维结构[1]。相比于依赖多张图像输入的多视图三维重建，单视图方法在输入数据获取便捷性和应用场景广泛性方面具有明显优势，特别适用于数据采集受限、目标对象无法多角度拍摄或仅有历史图像可用的场景，如历史建筑复原[2] [3]医学放射治疗[4]、自动驾驶车辆避障[5]、以及机器人导航[6] [7]等多个领域。然而，由于单视图图像仅包含有限的视角信息，缺乏立体几何约束，使得从单张图像中准确恢复完整三维结构仍面临极大挑战，因此单视图三维重建成为当前研究热点。

根据三维空间表示形式，现有的单视图三维重建方法主要分为隐式表达方法和显式表达方法[8]两类。基于隐式表达的单视图三维重建是通过函数来描述物体的几何形状来进行物体的三维重建[9]。隐式表达方法主要有占用场(Occupancy Field) [10]、符号距离函数(Signed Distance Function, SDF) [11]、神经辐射场(Neural Radiance Field, NeRF) [12]等。基于显式表达的单视图三维重建方法是指明确定义几何形状和结构以直接描述物体的外部或内部几何特征进行物体的三维重建。相对于隐式表达显示表达更易于可视化和编辑。显示表达方法主要有点云[13]、网格[14]以及 3DGS [15]表达三种方法。前者侧重函数建模，适合高保真重建，但对数据质量与推理效率要求高；后者具有结构可视性强、渲染速度快的优势。特别是该类方法中的 3D Gaussian splatting (3DGS)能够在消费级硬件上实现了高质量 3D 场景生成与实时渲染的有机结合。因此 3DGS 成为了当前单视图三维重建的热点。

然而 3DGS 其在复杂几何细节建模与点云分布适应性方面仍存在明显不足。具体而言，标准 3DGS 在处理物体边缘、遮挡区或纹理剧烈变化区域时，往往出现渲染模糊或结构失真；此外，其固定的高斯参数生成过程难以根据场景复杂度动态调整点密度与协方差结构，从而限制了其在单视图场景中的表达能力。为此，本文提出了一种名为 MMAGS 的高精度单视图三维重建框架，并将其应用于多类数据集中的单视图三维重建任务。该框架在图像精度、结构一致性与感知质量方面均展现出优越性能。主要贡献有以下三点：

- (1) 设计了一种由并行特征提取与三维高斯优化渲染的两阶段单视图三维重建框架，前阶段采用

Vision Mamba 模块并行提取全局特征，后阶段基于三维高斯优化的 Gaussian Splatting 实现高质量三维重建。该框架在 ShapeNet-SRN 和 CO3D 等数据集上表现出优异的跨类别泛化能力与复杂真实场景下的重建鲁棒性。

(2) 一种多通道 Vision Mamba 并行特征提取结构被构建，通过将中间特征划分为多个子通道并行输入多个 Mamba 模块，提升了跨视角结构信息的稳健提取能力。在遮挡区域展现出更强的几何还原效果，同时提升了模型泛化能力；

(3) 一种基于深度 - 颜色双梯度感知自适应 3D 高斯平滑滤波器被提出，可根据每个高斯点的局部几何变化与纹理复杂度动态调整其协方差，显著提升了重建图像的几何锐度与细节保真度，有效消除伪影抑制了边缘模糊与结构混叠现象。

## 2. 本文的方法

为实现高精度的单视图三维重建，本文提出 MMAGS 网络框架，如图 1 所示。该框架分为两个主要部分：(a) 多通道并行 Mamba 高斯预测网络(MPM Gaussian Prediction Network)。(b) 双梯度感知的 Gaussian Splatting 优化(Gradient-Aware GS Optimization)。在(a)中，输入的二维图像首先经过由  $3 \times 3$  卷积模块提取浅层特征，随后输入至由多个 Vision Mamba 层(VM Layer  $\times$  N)堆叠组成的子网络逐层建模跨区域结构依赖关系提取深层次特征。该结构在 VM 层后施加 LayerNorm 正则化，以稳定训练并增强特征一致性。最终  $1 \times 1$  卷积模块进行特征融合输出一个三维高斯参数图 G (即 MPM-image)，其中每个像素位置均包含一个三维高斯分布的参数集合(透明度、位置、颜色、协方差)。在(b)中，MPM-image 中编码的高斯参数被映射为具体的三维高斯点集合。随后，深度 - 颜色双梯度感知自适应 3D 高斯滤波(3D filter)依据高斯点所在区域的深度梯度与颜色梯度可对协方差进行有针对性的调整，从而在边缘与高频区域增强细节表达在平坦区域保持平滑连续。经过滤波后，生成的高斯点被用于执行可微喷溅渲染，最终生成结构一致、细节丰富的多视角重建结果。

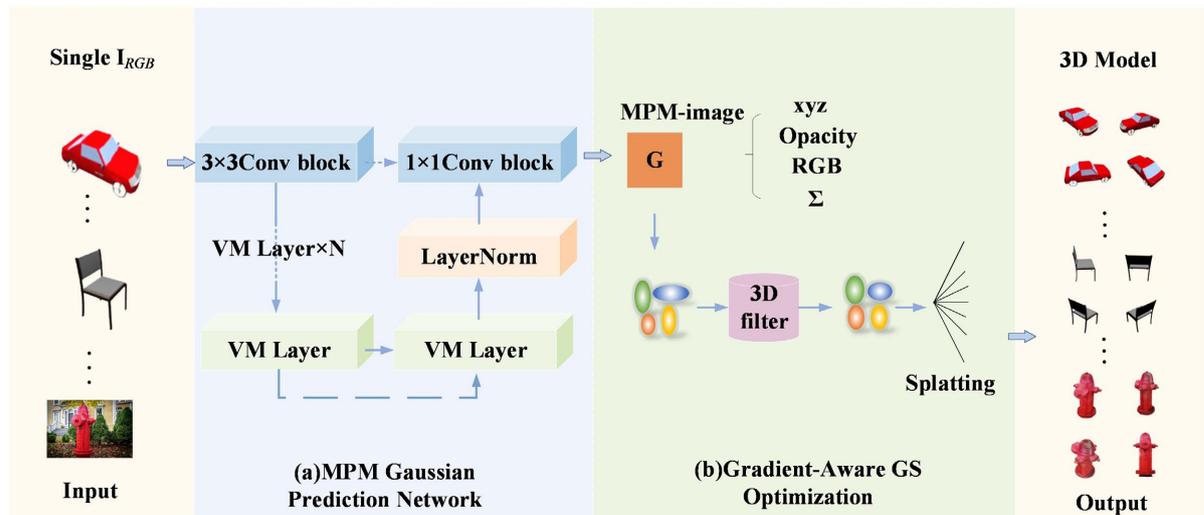


Figure 1. MMAGS network

图 1. MMAGS 网络框架

### 2.1. 多通道并行 Mamba 高斯预测网络

为了增强特征建模能力的同时保持推理效率，本文在特征提取阶段设计了一种多通道并行 Mamba 高

斯预测网络(MPM Gaussian Prediction Network)。其网络整体结构如图 2(a)所示, 该模块的核心是 VM layer 层见图 2(b), 该层中的 Mamba 结构见图 2(c)。本节将从整体架构设计、并行特征提取 VM layer 层、状态空间建模单元 Mamba 三部分详细介绍多通道并行 Mamba 高斯预测网络的三维高斯参数生成过程。

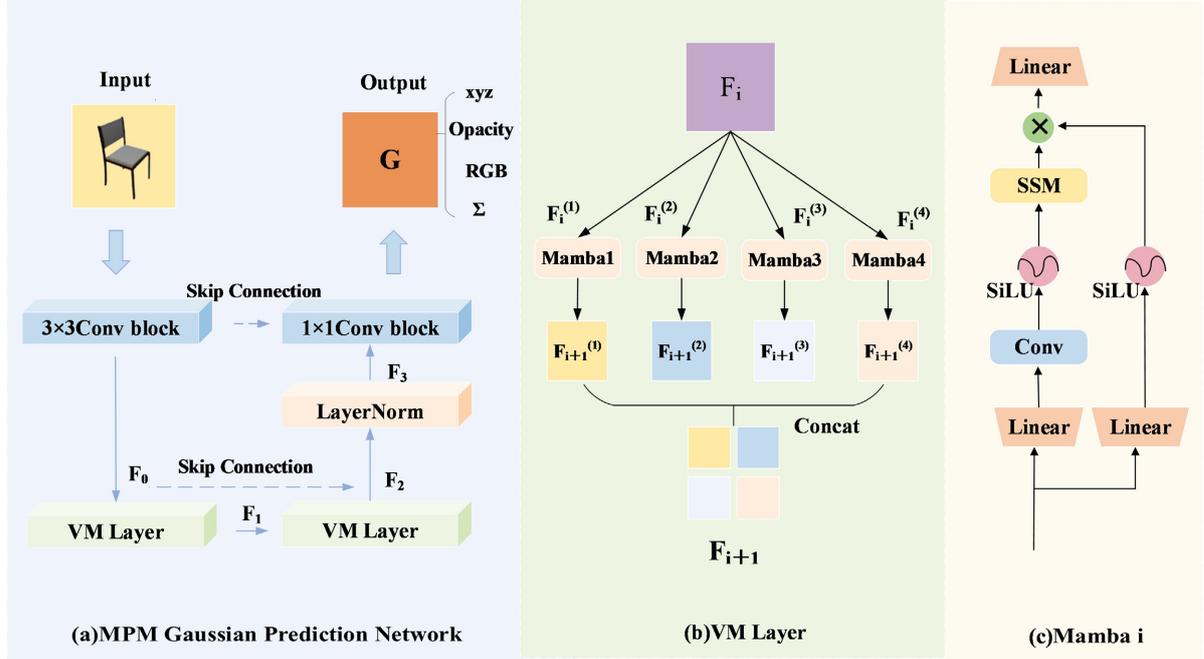


Figure 2. Multi-channel parallel vision Mamba gaussian prediction network  
图 2. 多通道并行 Mamba 高斯预测网络

(1) 整体架构设计

多通道并行 Mamba 高斯预测网络旨在实现局部特征提取与全局依赖关系统一建模整体结构如图 2(a)所示。

设输入图像为  $I \in \mathbf{R}^{H \times W \times 3}$ , 通道数为 3, 空间尺寸为  $H \times W$ 。网络最终输出为三维高斯参数图  $G \in \mathbf{R}^{H \times W \times K}$ , 其中  $K$  表示每个像素位置上预测的高斯参数维度(包括透明度、位置、颜色、协方差)。特征提取流程分为以下五个步骤:

① 初步特征提取:

输入图像  $I$  经过  $3 \times 3$  卷积模块  $\text{Conv3}()$  提取边缘与纹理特征, 生成初始中间特征图  $F_0 \in \mathbf{R}^{H \times W \times C}$ , 见公式(1)。

$$F_0 = \text{Conv3}(I) \tag{1}$$

② VM Layer 深层特征提取:

$F_0$  送入串联的两层 VM Layer 模块进行深层特征提取得到  $F_2$ , 具体实现细节见本节(2)并行特征提取 VM layer 层。

③ 残差融合与归一化:

VM Layer 的输出特征  $F_2$  与初始特征  $F_0$  进行跳跃连接并进行 Layer Normalization 得到融合特征  $F_3$ , 见公式(2)。

$$F_3 = \text{LayerNorm}(F_0 + F_2) \tag{2}$$

## ④ 三维高斯参数映射:

$1 \times 1$  卷积层  $\text{Conv1}()$  对  $F_3$  卷积, 通过跳跃链接进行特征融合并映射为三维高斯参数图  $G$ , 见公式(3)。

$$G = \text{Conv1}(F_3), G \in \mathbf{R}^{H \times W \times K} \quad (3)$$

该输出  $G$  中包含每个像素位置的高斯透明度  $\alpha$ 、三维位置  $(x, y, z)$ 、颜色(RGB)以及协方差矩阵参数  $\Sigma$ , 即为最终用于渲染的 MPM-image。

## (2) 并行特征提取 VM layer 层

本模块中的核心 VM layer 层见图 2(b)。其将输入特征图  $F_i$  (本文使用两次 VM layer 层故  $i=0,1$ )沿通道维度划分为  $n=4$  个子通道块  $F_i^{(1)}, F_i^{(2)}, F_i^{(3)}, F_i^{(4)}$  用于实现高效的特征提取, 选择  $n=4$  是基于 Vision Mamba 模块在处理单视图特征时能够充分利用 GPU 并行计算能力的同时, 确保各子通道间的特征不会过度分散而影响遮挡区域几何信息的有效建模, 见公式(4)。

$$F_i = (F_i^{(1)}, F_i^{(2)}, \dots, F_i^{(n)}), n \in [1, 2, 3, 4] \quad (4)$$

$F_i$  每个子通道块  $F_i^{(n)}$  被分别送入独立的 Mamba 子模块进行状态建模(具体实现细节见本节(3)状态空间建模单元 Mamba), 输出特征为  $F_{i+1}^{(n)}$ 。所有分支输出  $F_{i+1}^{(n)}$  拼接为完整特征  $F_{i+1}$ , 见公式(5)。

$$F_{i+1} = \text{Concat}(F_{i+1}^{(1)}, F_{i+1}^{(2)}, \dots, F_{i+1}^{(n)}), n \in [1, 2, 3, 4] \quad (5)$$

该结构允许在保持输入输出分辨率不变的同时, 进行高效的并行特征提取。

## (3) 状态空间建模单元 Mamba

Mamba 模块是 VM layer 中的基础构建单元, 用于高效建模特征中的长距离空间依赖。该模块整体结构包含线性变换、状态传播与输出映射三阶段[16] [17], 如图 2(c)所示。

具体而言, 输入特征  $F_i^j, i \in [1, 2], j \in [1, 4]$  经过两个并行的通道分别构建主干路径与状态路径。其中状态路径依次经过线性层 Linear、卷积 Conv、SiLU 激活与 Selective State Space Model (SSM)操作, 对特征进行动态状态更新; 主干路径则经过线性层 Linear、SiLU 激活保留输入中的局部响应信息。最终, 两条路径通过逐元素乘法进行融合, 整合结构感知能力与全局建模能力, 并由上层线性层完成输出映射。该结构在保持线性计算复杂度的同时, 有效增强了对长程依赖与跨区域结构的建模能力。

## 2.2. 基于双梯度感知的 Gaussian Splatting 优化

为了进一步提升三维重建的结构保真度与视觉一致性, 本文设计了一种基于深度 - 颜色双梯度感知自适应 3D 高斯滤波的 Gaussian Splatting 优化, 如图 3 所示。

## 2.2.1. 整体流程

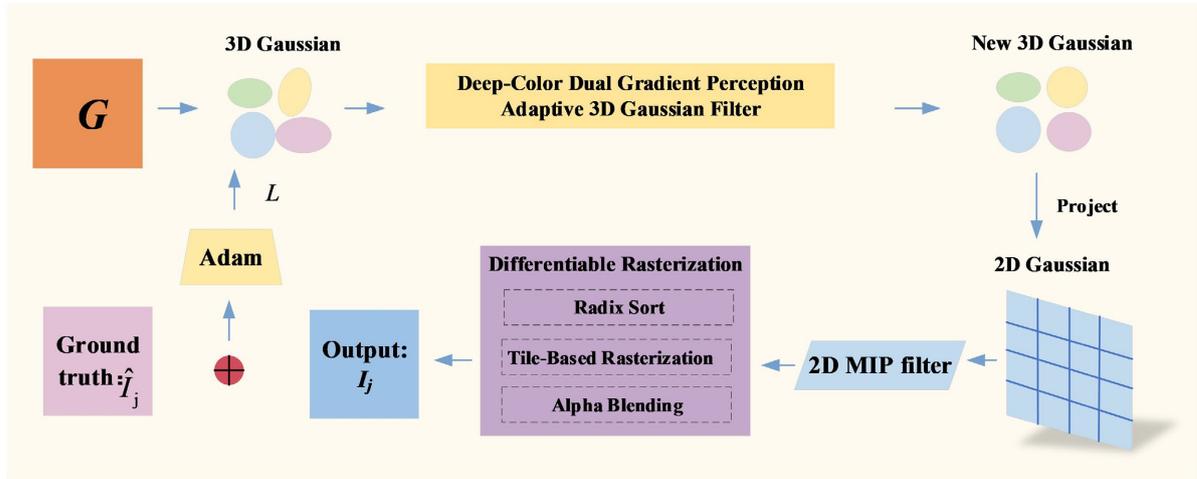
基于深度 - 颜色双梯度感知自适应 3D 高斯滤波的 Gaussian Splatting 优化整个优化流程可分为以下六个步骤:

## (1) 3D 高斯生成

在 MPM Gaussian Prediction Network 生成高斯参数图  $G \in \mathbf{R}^{H \times W \times K}$  的每个像素位置  $(u_i, v_i) \in [0, H-1] \times [0, W-1]$  都对对应生成一个高斯原语  $\mathcal{G}_i$ , 其参数见公式(6)。

$$\mathcal{G}_i = \{x_i, \Sigma_i, c_i, \alpha_i\} \quad (6)$$

其中,  $x_i \in \mathbf{R}^3$  为高斯中心位置,  $\Sigma_i \in \mathbf{R}^{3 \times 3}$  为协方差矩阵,  $c_i \in \mathbf{R}^3$  为颜色向量,  $\alpha_i \in [0, 1]$  为透明度(opacity)。



**Figure 3.** Gradient- and density-aware Gaussian Splatting optimization  
**图 3.** 双梯度感知的 Gaussian Splatting 优化流程

(2) 3D 高斯自适应优化

深度 - 颜色双梯度感知自适应 3D 高斯滤波对每个 3D 高斯原语  $G_i$  进行低通滤波处理, 目的是限制高频分量, 防止图像生成时出现伪影。即对于每个高斯原语  $G_i$ , 根据其深度梯度  $c_i^{depth} \in [0,1]$  与颜色梯度  $c_i^{color} \in [0,1]$  动态调节其协方差矩阵(详见 2.2.2)。

(3) 三维投影至二维平面

优化后的 3D 高斯分布通过模拟相机投影到 2D 图像平面。每个高斯点  $i$  的中心  $x_i$  投影得到图像像素坐标  $(u'_i, v'_i) \in [0, W - 1] \times [0, H - 1]$  见公式(7)。

$$(u'_i, v'_i) = Project(x_i, Camera) \tag{7}$$

投影后的位置用于确定在屏幕上的影响范围和渲染排序。

(4) 2D Mip 滤波器抗混叠

缓解因视角缩放或远距离投射导致的高斯混叠伪影, 本文在投影阶段引入了 2D Mip 滤波器[18]。该滤波器为每个高斯点引入固定尺度的模糊核, 在局部区域内对像素值执行加权平均, 使中心区域响应更强、边缘逐渐衰减, 从而在保持细节的同时实现平滑过渡。这一处理策略不仅抑制了混叠伪影, 也提升了图像边缘的一致性与整体清晰度。

(5) 快速可微光栅化

① 深度排序(Radix Sort): 按照每个高斯原语  $G_i$  的视角深度对平滑处理后的二维高斯使用 GPU 基数排序(Radix Sort)进行排序, 确保渲染过程中前后顺序正确, 从而实现合理的透明度叠加。

② 切片化光栅化(Tile-Based Rasterization): 屏幕划分为固定大小的瓦片(如  $16 \times 16$  像素), 逐切片渲染其中可见的高斯点, 提高并行计算效率。

③ 透明度混合(Alpha Blending): 在同一像素上多个高斯点的颜色和透明度进行前向累积叠加, 其目标是合成自然、渐变的渲染图像, 同时保持可微性以支持反向传播。

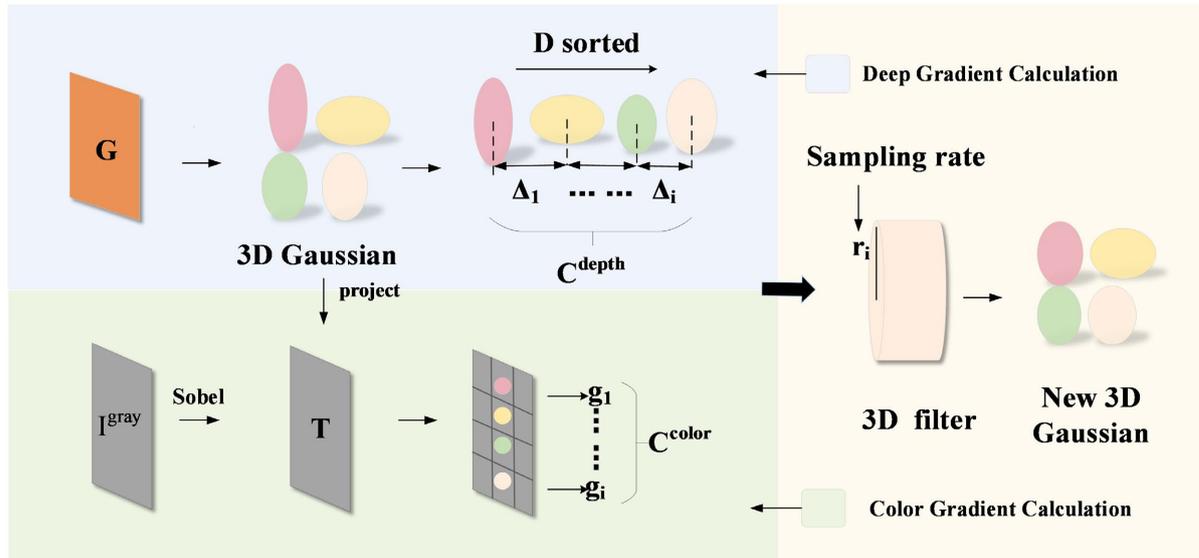
(6) 参数优化

在渲染过程中不断使用 Adam 优化器最小化高斯渲染图像  $I_j$  与真实图像  $\hat{I}_j$  在训练视角  $j$  下的多项损失函数  $L$ , 见公式(8)。该损失函数综合考虑了图像精度(PSNR, 峰值信噪比)、结构相似性(SSIM, 结构相似性指数)和感知一致性(LPIPS, 感知相似度)三方面构成,  $\lambda_1$ 、 $\lambda_2$ 、 $\lambda_3$  为对应的系数。

$$L = \sum_j \left[ \lambda_1 \cdot \text{PSNR}(I_j, \hat{I}_j) + \lambda_2 \cdot \text{SSIM}(I_j, \hat{I}_j) + \lambda_3 \cdot \text{LPIPS}(I_j, \hat{I}_j) \right] \quad (8)$$

损失通过反向传播同时优化所有可学习高斯参数，包括每个高斯的空间位置、颜色、透明度以及协方差矩阵。优化过程持续迭代，直到达到预设的训练轮次或损失收敛为止。该过程如图 3 中所示，输出图像  $I_j$  与对应真实图像  $\hat{I}_j$  比较生成损失  $\mathcal{L}$ ，再通过 Adam 优化器将梯度反向传递至高斯点参数，驱动三维重建不断收敛优化。

### 2.2.2. 深度 - 颜色双梯度感知自适应 3D 高斯滤波



**Figure 4.** Depth-color dual-gradient perception adaptive 3D gaussian filtering  
**图 4.** 深度 - 颜色双梯度感知自适应 3D 高斯滤波

在 Mip-Splatting [18] 的原始方法中，为避免伪影现象，每个 3D 高斯原语的滤波半径  $r_i$  根据该点在各训练视角中所能达到的最大采样频率进行计算。其采样频率受相机焦距  $f$  与点深度  $d_i$  的比值控制，计算过程可简化表示为公式(9)。

$$r_i = \frac{d_i}{f} \cdot \sqrt{0.2} \quad (9)$$

该公式仅考虑几何关系未引入图像内容复杂度，导致以下两类问题：在纹理丰富或边缘区域，过度的平滑操作会模糊细节降低渲染保真度；而在结构简单、纹理均匀的区域，统一的滤波尺度又可能不足以抑制噪声或伪影。

为此，本文提出一种深度 - 颜色双梯度感知自适应 3D 高斯滤波，设计内容引导的滤波尺度调整机制动态平衡抗混叠与细节保留，如图 4 所示。该方法是基于深度梯度和颜色梯度来估计每个高斯点的场景复杂度实现平滑过程的优化。

#### (1) 深度梯度计算

深度梯度是指高斯点在空间中的深度变化，其本质反映了该点所处位置在三维空间中的深度突变程度。深度梯度计算可刻画高斯点是否处于深度突变区域，例如物体边缘、遮挡边界等，该计算过程主要有以下 6 个步骤：

① 设输入图像中共有  $N$  个高斯点，点的深度  $d_i$  集合可记作  $D$ ，见公式(10)。

$$D = \{d_1, d_2, \dots, d_N\}, d_i \in R \quad (10)$$

② 所有深度值  $D$  进行升序排序得到  $D_{\text{sorted}}$ ，见公式(11)。

$$D_{\text{sorted}} = \text{sort}(D) \quad (11)$$

③ 相邻点之间的深度差值  $\Delta_i$  (近似一维梯度)被计算,表示每对相邻高斯点的深度变化,见公式(12)。

$$\Delta_i = |D_{\text{sorted}}[i+1] - D_{\text{sorted}}[i]|, i \in [1, N-1] \quad (12)$$

④ 深度差值集合被扩展为与原深度集合  $D$  长度一致,在集合末尾增加一个初始值为  $\Delta_{N-1}$  的深度差值即  $\Delta = \{\Delta_1, \Delta_2, \dots, \Delta_{N-1}, \Delta_{N-1}\}$ 。

⑤ 该差值集合  $\Delta$  按索引映射回原始点集合,计算深度差值集合  $\Delta$  中每个深度差值  $\Delta_i$  减去集合中的最小值  $\min(\Delta)$ ,再除以集合中的最大值  $\max(\Delta)$  减去最小值  $\min(\Delta)$  的差值进行归一化,见公式(13)。

$$c_i^{\text{depth}} = \frac{\Delta_i - \min(\Delta)}{\max(\Delta) - \min(\Delta)} \quad (13)$$

⑥ 归一化后的几何深度梯度  $c_i^{\text{depth}}$  构成深度梯度集合  $C^{\text{depth}}$ ，见公式(14)。

$$C^{\text{depth}} = \{c_1^{\text{depth}}, c_2^{\text{depth}}, \dots, c_N^{\text{depth}}\}, c_i^{\text{depth}} \in [0, 1] \quad (14)$$

## (2) 颜色梯度计算

颜色梯度用于衡量图像中颜色或亮度的局部变化强度。颜色梯度计算有效反映高斯点是否处于颜色纹理变化剧烈区域,如物体轮廓、纹理边界等,该方法主要有以下7个步骤。

① 设每个训练图像为  $I \in R^{H \times W \times 3}$ , 图像的红色、绿色和蓝色通道的像素值按一定权重进行加权平均得到的其对应的灰度图  $I^{\text{gray}}$  见公式(15)。

$$I^{\text{gray}} = 0.299 \cdot I(x, y)^{(R)} + 0.587 \cdot I(x, y)^{(G)} + 0.114 \cdot I(x, y)^{(B)} \quad (15)$$

② Sobel [19]算子分别计算图像  $I^{\text{gray}}$  的横向梯度  $T_x$  和纵向梯度  $T_y$ ，见公式(16)

$$\begin{cases} T_x = I^{\text{gray}} * K_x \\ T_y = I^{\text{gray}} * K_y \end{cases} \quad (16)$$

其中  $K_x$ 、 $K_y$  别为横向、纵向的 Sobel 卷积核。

③ 根据横向梯度  $T_x$  和纵向梯度  $T_y$ ，每个像素  $I(x, y)$  的梯度幅值被计算得到颜色梯度图  $T$ ，见公式(17)。

$$T = \sqrt{T_x^2 + T_y^2} \quad (17)$$

④ 每个高斯点  $i$  投影到图像平面  $(u_i, v_i)$ ，并在图像梯度图  $T$  上用双线性插值采样获取该点在当前视角的颜色梯度值  $g_i$ ，见公式(18)。

$$g_i = T(u_i, v_i) \quad (18)$$

⑤ 高斯点  $i$  在所有视角中采样到的最大颜色梯度值  $\max_{j \in \text{views}}(g_i^j)$  保留作为该点最终颜色梯度指标,并由此归一化得到颜色梯度  $c_i^{\text{color}}$ ，见公式(19)。

$$c_i^{\text{color}} = \frac{\max_{j \in \text{views}}(g_i^j) - \min(g)}{\max(g) - \min(g)} \quad (19)$$

⑥  $c_i^{\text{color}}$  合并成颜色梯度集合  $C^{\text{color}}$ , 见公式(20)。

$$C^{\text{color}} = \{c_1^{\text{color}}, c_2^{\text{color}}, \dots, c_N^{\text{color}}\}, c_i^{\text{color}} \in [0, 1] \quad (20)$$

(3) 基于深度与颜色梯度的自适应 3D 平滑滤波器

在获得每个高斯点的归一化深度梯度  $c_i^{\text{depth}} \in [0, 1]$  和颜色梯度  $c_i^{\text{color}} \in [0, 1]$  后, 其作为场景复杂度的指标引入滤波半径的调节中, 构建了一种基于内容感知的自适应 3D 滤波机制。其主要分为滤波半径设置、高斯分布和自适应滤波三个部分。

① 滤波半径设置

对原始滤波半径进行深度梯度和颜色梯度联合调节, 构造新的滤波半径  $r_i$  计算见公式(21)。

$$r_i = \left( \frac{d_i}{f} \cdot \sqrt{0.2} \right) \cdot (1 - \lambda_d \cdot c_i^{\text{depth}} - \lambda_c \cdot c_i^{\text{color}}) \quad (21)$$

其中,  $\lambda_c$ 、 $\lambda_d \in [0, 1]$  是调节超参数(本工作中根据经验性设为 0.4, 该数值确保深度梯度和颜色梯度对滤波半径的调节作用适中, 既能在高变化区域有效缩小滤波半径以保留细节, 又避免过度调节导致的渲染不稳定),  $0 \leq \lambda_c + \lambda_d \leq 1$ 。当  $c_i^{\text{depth}}$  和  $c_i^{\text{color}}$  较大时, 表示该点处存在显著的深度变化和颜色变化, 此时缩小滤波半径以保留更多细节; 反之在结构平坦区域, 滤波半径则保持较大, 有助于抑制噪声与伪影现象。

② 高斯分布

为实现实际滤波效果, 为每个高斯点引入一个深度 - 颜色双梯度感知的滤波半径  $r_i$ 。在训练阶段, 原始 3D 高斯  $G_i(x) \sim \mathcal{N}(\mu_i, \Sigma_i)$  在其中心位置处被施加以  $r_i$  为带宽的高斯核进行平滑操作, 所得到的新高斯分布的协方差见公式(22)。

$$\Sigma_i^{\text{filtered}} = \Sigma_i + r_i^2 \cdot I \quad (22)$$

其中,  $I$  为三维单位矩阵, 这一结果利用了高斯核的卷积封闭性, 即两个高斯分布卷积后仍为高斯, 其均值相加, 协方差矩阵相加。

③ 自适应滤波

在完成协方差更新后, 每个高斯点  $G_i^{\text{filtered}}(x)$  在空间中新的表达式可见公式(23)。

$$G_i^{\text{filtered}}(x) = \sqrt{\frac{|\Sigma_i|}{|\Sigma_i^{\text{filtered}}|}} \cdot \exp\left(-\frac{1}{2}(x - \mu_i)^T (\Sigma_i^{\text{filtered}})^{-1} (x - \mu_i)\right) \quad (23)$$

其中,  $\mu_i$  为高斯点中心, 该滤波过程在训练阶段对每个高斯点独立执行, 其结果直接参与协方差优化, 并成为场景表示的固有部分, 训练后不再随视角变化而改变。该基于深度梯度和颜色梯度的自适应 3D 平滑滤波器, 不仅保留了 3DGS 的渲染效率与结构表达能力, 同时大幅提升了模型对视角变化和分辨率变化的稳健性。

### 3. 实验结果与分析

#### 3.1. 数据集

为了全面评估本文方法在不同场景下的重建性能, 本文在标准数据集 ShapeNet-SRN 与真实数据集 CO3D 上分别开展实验, 涵盖结构复杂度与纹理精细度各异的典型类别。

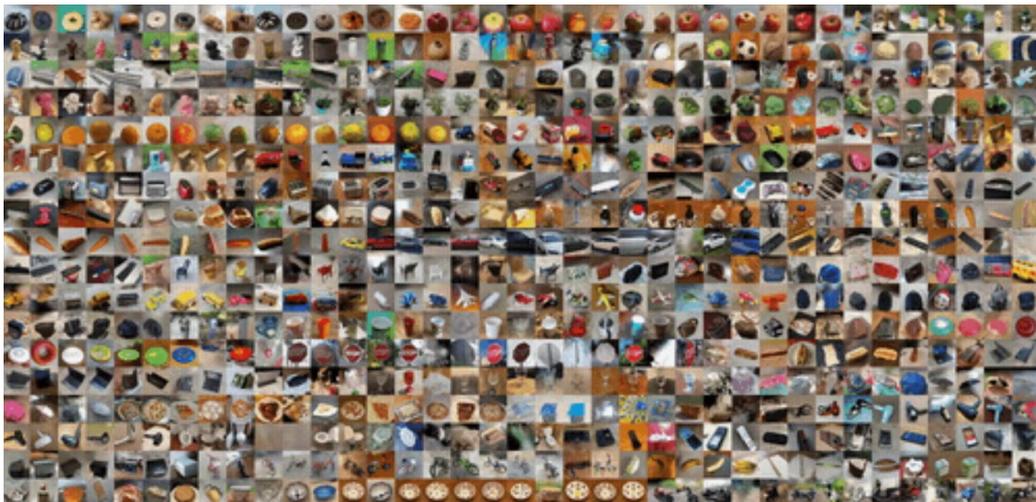
ShapeNet-SRN [20] [21] 是评估单视角 3D 重建的标准基准数据集, 是从 ShapeNet v2 数据集中挑选并使用 Blender 重新渲染成统一格式的子集。该数据集中只包含 “Cars”、“Chairs” 两个类见图 5 所示, 其中 Cars 包含了 3514 个实例、Chairs 包含 689 个实例、每个实例包含 100 张分辨率为  $128 \times 128$  RGB 渲

染图像、每张图像对应的相机姿态(位姿)、用于训练的相机内外参。



**Figure 5.** ShapeNet-SRN dataset  
**图 5.** ShapeNet-SRN 数据集

CO3D [22] [23]是由 Facebook AI Research 发布的大规模真实三维重建数据集,包含约 1.9 万个视频,覆盖 50 个常见类别(如苹果、椅子、鞋等)如图 6 所示,总计约 140 万帧 RGB 图像。每个物体实例都配有相机内外参、2D 分割掩码和类别标签。数据集由真实设备拍摄,涵盖自然光照、背景干扰,更贴近实际应用场景,广泛用于训练 3D 模型。CO3D 数据集的许多图像中物体存在部分遮挡或裁剪,同时相机焦距等内参在不同序列之间存在显著变化,是具有挑战性的场景,适合测试算法的鲁棒性。



**Figure 6.** CO3D dataset  
**图 6.** CO3D 数据集

### 3.2. 评估指标

为全面评估所提方法在单视图三维重建任务中的性能表现,本文选择 PSNR (Peak Signal-to-Noise Ratio, 峰值信噪比)、SSIM (Structural Similarity Index, 结构相似性指数)、LPIPS (Learned Perceptual Image Patch Similarity, 感知相似度)三个指标[12] [15] [24],对生成三维模型图像与参考图像之间的差异进行定量分析。

(1) PSNR 是衡量图像质量的经典指标, 通常用于比较重建图像与原始图像之间的差异。在三维重建任务中, PSNR 用于评估从生成的三维模型得到的图像与参考图像之间的像素差异。PSNR 值越高, 表示重建图像与真实图像之间的差异越小, 图像质量越好。PSNR 通过计算原始图像中像素的最大可能值  $MAX_I$  (例如, 对于 8 位图像,  $MAX_I = 255$ ) 与重建图像之间的均方误差 (Mean Squared Error, MSE) [24] 来量化它们之间的差异, 见公式(24)。

$$PSNR = 10 \cdot \log_{10} \left( \frac{MAX_I^2}{MSE} \right) \quad (24)$$

其中 MSE 计算原图像的像素值  $I(i)$  和重建图像  $K(i)$  之间的误差, 具体公式见(25)。

$$MSE = \frac{1}{N} \sum_{i=1}^N (I(i) - K(i))^2 \quad (25)$$

其中,  $N$  是图像的总像素数。

(2) SSIM 是一种用于衡量图像结构相似性的指标, 它考虑了图像的亮度、对比度和结构等因素。在三维重建中, SSIM 用于比较通过三维模型渲染得到的图像与参考图像之间的结构差异。SSIM 的值介于 -1 和 1 之间, 值越接近 1 表示两幅图像越相似。其具体计算见公式(26)。

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (26)$$

其中,  $\mu_x$  和  $\mu_y$  代表局部区域的平均亮度,  $\sigma_x^2$  和  $\sigma_y^2$  是方差,  $\sigma_{xy}$  是协方差。  $C_1 = (K_1L)^2$  和  $C_2 = (K_2L)^2$  是常数为了避分母为 0,  $L$  是图像的动态范围(例如, 对于 8 位图像,  $L = 255$ , 而  $K_1$  和  $K_2$  是小常数, 通常设置为  $K_1 = 0.01$  和  $K_2 = 0.03$ )。

(3) LPIPS 是一种基于深度学习的感知图像相似性度量, 它通过一个训练过的深度网络来衡量两张图像在感知上的差异。在三维重建中, LPIPS 通过计算渲染图像与真实图像之间的深层次特征差异来评估图像质量, 从而间接评估三维模型的质量。LPIPS 通过考虑高层次特征(如纹理和边缘信息)来评估图像的相似性, 因此能捕捉到人眼对细节的敏感度。

LPIPS 的计算是通过深度神经网络提取图像的特征表示后, 对比这两张图像的特征差异。首先两张图像  $I$  和  $J$  的特征  $\phi(I)$  和  $\phi(J)$  被提取, 然后网络总层数为  $N$  的第  $i$  层的特征表示  $\phi_i(I)$  和  $\phi_i(J)$  之间的 L2 距离被计算, 具体计算见公式(27)。

$$LPIPS(I, J) = \frac{1}{N} \sum_{i=1}^N \|\phi_i(I) - \phi_i(J)\|_2 \quad (27)$$

### 3.3. 实验配置

环境: 本实验中, 硬件环境使用了 CPU 型号为英特尔 13 代酷睿 i9-13900 处理器, 64 GB 内存, NVIDIA GeForce RTX 4080 显卡, 显存大小为 16 GB。操作系统使用了 ubuntu18.04.6 LTS 版本, python 版本为 3.8.19, pytorch 版本为 1.13.0, cuda 版本为 11.6。为适配大规模训练数据与结果存储, 实验环境配备了一块容量为 4 TB 的 NVMe 高速移动固态硬盘, 用于数据缓存与模型中间结果的快速读写。

训练参数配置: 训练不同的数据集时采用不同的训练参数, 并且将单视图模型训练分为两阶段。以 ShapeNet-SRN 中 cars 数据集为例, 将 batch size 设置为 8, 使用  $5 \times 10^{-5}$  的学习率, 在训练达到 800000 次迭代后, 进行第二阶段训练, 学习率减少 10 倍并使用 LPIPS 进行微调, 再继续训练迭代 100000 次。

### 3.4. 性能比较与分析

#### 3.4.1. 可视化结果比较

##### (1) 基于 ShapeNet-SRN 数据集的可视化结果比较

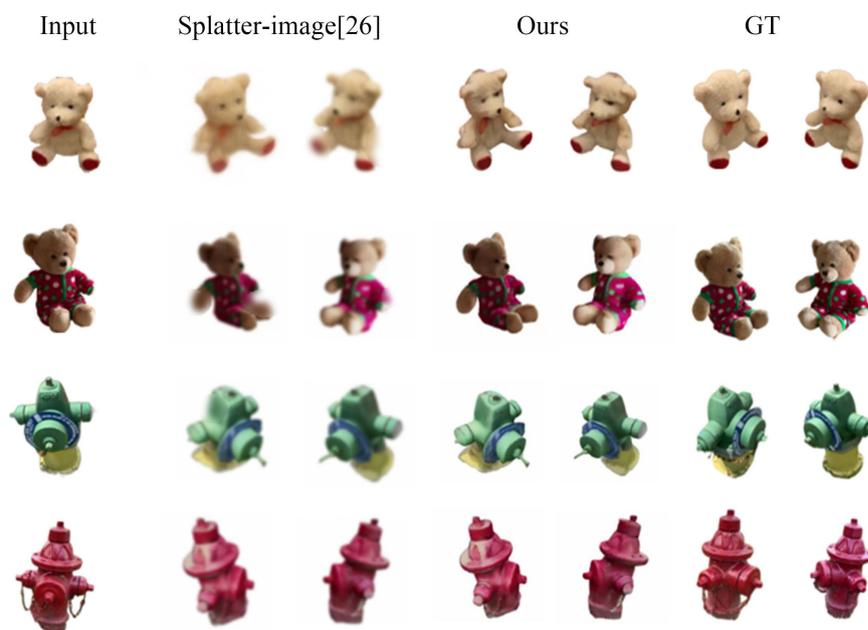


**Figure 7.** Visualization results comparison based on the ShapeNet-SRN dataset  
**图 7.** 基于 ShapeNet-SRN 数据集的可视化结果比较

在标准数据集 ShapeNet-SRN 上, 本文选取典型类别样本并对比 PixelNeRF [25]、VisionNeRF [26]、Splatter-image [24] 及本文提出的 MMAGS 方法的重建结果, 可视化结果如图 7 所示。其中每行对应一个具体类别(包括汽车与椅子)样本, 每列依次展示输入图像(Input)、主流方法的渲染结果(PixelNeRF、VisionNeRF、Splatter-image)、本文方法(Ours)以及参考视图(GT)。为便于观察重建细节, 图中对关键区域进行了局部标注。

在 Cars 类别中, MMAGS 在轮廓清晰度、结构一致性和高频细节还原方面表现出明显优势, 尤其在车头边缘、窗框遮挡及轮胎区域均重建完整, 几何形态自然准确。对于 Chairs 类别, MMAGS 同样展现出良好的结构表达能力, 椅背边缘曲线规整、空洞连接自然, 椅腿结构清晰分离, 空间感强。综上, 标准数据集上 ShapeNet-SRN 的可视化结果充分证明了本文方法在单视图三维重建任务中的有效性和优越性。无论是规则结构(如 Cars)还是形态复杂、连接精细的对象(如 Chairs), MMAGS 均能准确还原物体轮廓、边缘细节与结构空隙, 整体重建质量显著优于现有方法, 充分验证了其在标准数据集上的通用性与稳定性。

##### (2) 基于 CO3D 数据集的可视化结果比较



**Figure 8.** Visualization results comparison based on the CO3D dataset  
**图 8.** 基于 CO3D 数据集的可视化结果比较

为验证所提方法在真实复杂场景下的表现，在 CO3D 数据集中选取了两个典型类别(teddybear 与 hydrant)进行可视化对比，结果如图 8 所示。每类物体展示两个样本(上下两行)，分别比较了 Splatter-image 与本文方法 MMAGS 的重建效果，并与输入视图和真实图像(GT)进行对照。

在 teddybear 类别中，MMAGS 能准确重建耳部、四肢与服装等关键结构，图案纹理清晰，形体层次感良好。在 hydrant 类别中，模型成功还原了螺纹棱线与出水口结构，几何细节完整，未出现明显粘连或扭曲。这得益于 MMAGS 中 MPM Gaussian Prediction Network 对全局形体结构的建模能力，以及滤波器对高频边缘与空间过渡区域的联合优化，整体提升了复杂结构下的几何稳定性与渲染质量。综上，本文所提出的 MMAGS 模型在 CO3D 数据集的多个真实场景中均展现出优于现有方法的可视化重建质量，特别在轮廓边界清晰度、细节还原完整性与结构一致性方面表现突出。可视化结果进一步印证了 MMAGS 在处理真实图像中遮挡、模糊、复杂纹理等干扰因素下的强鲁棒性与泛化能力。

### 3.4.2. 量化指标的比较与分析

为了客观比较分析本文提出的 MMAGS 方法在单视图三维重建任务中的性能，本节将基于不同数据集进行定量指标对比，分析所提方法在多个数据集上的重建质量。

#### (1) 基于 SRN-Cars 标准数据集的量化指标评价

首先在 ShapeNet-SRN 数据集中的“Cars”类别上，对本文方法与当前主流单视图三维重建方法进行了定量性能比较，结果如表 1 所示。从表中可以看出，MMAGS 在 PSNR 和 SSIM 上均取得领先(PSNR = 25.78, SSIM = 0.95)，分别优于 Splatter-image 与 VisionNeRF，表明其在图像细节还原与结构一致性方面具备显著优势。尽管在 LPIPS 指标上与 Splatter-image 相近，MMAGS 仍显著优于 PixelNeRF 与 CodeNeRF，体现出更强的感知一致性。这些改进主要得益于 MMAGS 中 MPM Gaussian Prediction Network 中 Vision Mamba 模块对全局结构的稳定建模能力，使模型在重建过程中能够准确恢复物体框架与遮挡边界，显著提升了结构相似性(SSIM)和细节还原能力。

**Table 1.** Comparison of metrics for different methods on the SRN-Cars dataset  
**表 1.** 基于 SRN-Cars 数据集不同方法指标对比

Method	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
SRN [21]	22.25	0.88	0.129
FE-NVS [27]	22.83	0.91	0.099
CodeNeRF [28]	23.80	0.91	0.128
PixelNeRF [25]	23.17	0.89	0.146
VisionNeRF [26]	22.88	0.90	0.084
ViewsetDiff w/o [29]	23.21	0.90	0.116
NeRFDiff w/o [30]	23.95	0.92	0.092
Splatter-image [24]	24.00	0.92	<b>0.078</b>
MMAGS (Ours)	<b>25.78</b>	<b>0.95</b>	<b>0.079</b>

(2) 基于 SRN-Chairs 标准数据集的量化指标评价

**Table 2.** Comparison of metrics for different methods on the SRN-Chairs dataset  
**表 2.** 基于 SRN-Chairs 数据集的不同方法指标对比

Method	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
SRN [21]	22.89	0.89	0.104
FE-NVS [27]	23.21	0.92	0.077
CodeNeRF [28]	23.66	0.90	0.166
PixelNeRF [25]	23.72	0.90	0.128
VisionNeRF [26]	24.48	0.92	0.077
ViewsetDiff w/o [29]	23.21	0.90	0.116
NeRFDiff w/o [30]	24.80	<b>0.93</b>	0.070
Splatter-image [24]	24.43	<b>0.93</b>	0.067
MMAGS (Ours)	<b>25.93</b>	<b>0.93</b>	<b>0.066</b>

在 SRN 数据集的“Chairs”类别中，本文方法同样表现出显著优势，如表 2 所示。从表中可以看出，本文的方法 MMAGS 其在 PSNR 上取得 25.93，显著优于其他方法，LPIPS 为 0.066，为所有方法中最低，表明在图像还原与感知质量方面具备双重优势。尽管 SSIM 与部分方法持平，MMAGS 在复杂几何结构（如椅背空隙、细长支撑）下仍能保持较高的结构一致性，有效缓解粘连与模糊现象。本文所提方法 MMAGS 在该类别表现优异，主要得益于其引入的自适应 3D 平滑滤波器机制。该模块可根据边缘梯度调节高斯协方差，在空洞区域保留结构锐度，在平坦区域避免过度模糊，有效提升了细节表达的稳定性与 PSNR 指标表现，并在 LPIPS 上保持较低感知误差。

(3) 基于 CO3D-hydrant 真实数据集的量化指标评价

在 CO3D-hydrant 真实数据集上，本文所提出的 MMAGS 与当前主流方法进行了定量性能比较，结果见表 3。该数据集包含复杂的光照、遮挡与形态变化，对模型的鲁棒性与泛化能力提出了更高要求。MMAGS 在 PSNR、SSIM 和 LPIPS 三项指标上均优于 Splatter-image 与 PixelNeRF，展现出更强的图像还原能力、结构一致性与感知质量。分析原因可知，该类样本结构复杂，包含多个曲面连接与细节棱角。

MMAGS 在该场景下取得优势，主要由自适应 3D 平滑滤波器在高频区域增强了边缘表达能力显著抑制了模糊与纹理丢失等常见问题，支撑 SSIM 和 LPIPS 的双重表现，同时避免了在背景区域出现冗余密度和感知失真。

**Table 3.** Comparison of metrics for different methods on the CO3D-hydrant dataset

**表 3.** 基于 CO3D-hydrant 数据集的不同方法指标对比

Method	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
PixelNeRF [25]	21.76	0.78	0.203
Splatter-image [24]	21.80	0.80	0.150
MMAGS (Ours)	<b>22.69</b>	<b>0.83</b>	<b>0.112</b>

(4) 基于 CO3D-Teddybear 真实数据集的量化指标评价

**Table 4.** Comparison of metrics for different methods on the CO3D-Teddybear dataset

**表 4.** 基于 CO3D-Teddybear 数据集的不同方法指标对比

Method	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
PixelNeRF [25]	19.38	0.65	0.290
Splatter-image [24]	19.44	0.73	0.150
MMAGS (Ours)	<b>20.19</b>	<b>0.75</b>	<b>0.123</b>

在 CO3D-Teddybear 真实世界数据集上，对 MMAGS 与当前主流单方法进行了量化性能对比，结果如表 4 所示。该类样本具有毛绒材质、纹理密集、边界柔和等特征，对细节保真与感知一致性构成严峻挑战。通过以上三项指标的对比可以看出，MMAGS 在 CO3D-Teddybear 数据集上表现出全面领先的性能。本文方法 MMAGS 在该场景下的表现得益于 Vision Mamba 模块与自适应 2D 平滑滤波器之间的协同作用：前者增强了毛绒物体形体结构的语义建模能力，后者则对边缘与高频纹理区域进行了密度强化，使模型在感知自然度与结构连续性之间实现有效平衡，显著提升了 PSNR 与 LPIPS 指标表现。这些模块共同作用使 MMAGS 能在高难度真实场景下实现高质量、结构稳定的三维重建结果，再一次证明本文方法在真实数据集上具有强鲁棒性与更好的泛化能力。

### 3.4.3. 消融实验

为了进一步验证本文所提出各关键模块对整体性能的贡献以及本文设计的方法组合达到最优，本节在 SRN-Cars 数据集上进行了消融实验。逐步增加多通道并行 Mamba 高斯预测网络(表 5 中 MPM)、深度 - 颜色双梯度感知自适应 3D 高斯滤波(表 5 中 Adaptive 3D filter)两个模块,比较不同组合下的重建效果,结果见表 5。从表中可以看出,当两个模块全部移除时,模型在 PSNR、SSIM 和 LPIPS 三项指标上分别为 24.00、0.92 和 0.078,性能处于最低水平。仅设计多通道并行 Mamba 高斯预测网络后,PSNR 提升至 25.00,说明该模块能够显著增强全局结构感知能力,提高像素级还原精度;但同时 SSIM 无变化,说明仅依赖全局建模尚不足以改善结构一致性。在此基础上进一步引入自适应 3D 平滑滤波器后 PSNR 上升至 25.78,SSIM 提升至 0.95,但 LPIPS 反而略升至 0.079,说明该模块有效提升了高频区域的细节表达与几何精度,但也可能引入边缘区域的过度锐化,从而在感知空间中略微降低了图像的自然性与连续性。MMAGS 框架中的这两个核心模块在结构保真、感知质量与重建精度方面均具有明确且互补的性能增益,验证了其设计的合理性与实用性。

**Table 5.** Ablation experiment on SRN-Cars dataset  
**表 5.** SRN-Cars 数据集上的消融实验

MPM	Adaptive 3D filter	PSNR ↑	SSIM ↑	LPIPS ↓
×	×	24.00	0.92	0.078
√	×	25.00	0.92	0.085
√	√	25.78	0.95	0.079

#### 4. 结论

在本研究中，我们提出了一种多通道 Mamba 自适应高斯溅射网络(MMAGS)，以应对单视图三维重建中精度受限与泛化能力不足的问题。该方法引入多通道并行 Vision Mamba 结构，通过特征通道划分与并行建模，有效增强了遮挡区域的几何还原能力与跨视角结构信息的稳定提取。同时，提出基于深度与颜色双梯度感知的自适应三维高斯滤波策略，能够根据局部几何变化与纹理复杂度动态调整高斯协方差。本文方法 MMAGS 在标准数据集 ShapeNet-SRN 和具有挑战的真实数据集 CO3D 上进行了测试实验。实验结果表明，本文方法在 PSNR、SSIM 和 LPIPS 三项核心指标上均显著优于现有先进的单视图三维重建方法。在 Cars 与 Chairs 等结构特征差异明显的类别中，MMAGS 有效缓解了边缘模糊、结构粘连与空洞闭合等问题，展现出稳定的几何还原能力；在真实场景下的 hydrant 与 teddybear 等类别中，MMAGS 同样表现出良好的感知一致性与鲁棒性，在复杂结构、遮挡区域与模糊纹理恢复方面均取得显著效果。

#### 参考文献

- [1] Belkaid, M., Alaoui, E.A.A., Berrajaa, A., Akkad, N.E. and Merras, M. (2024) Deep Learning-Based Solution for 3D Reconstruction from Single RGB Images. 2024 *International Conference on Circuit, Systems and Communication (IC-CSC)*, Fes, 28-29 June 2024, 1-6. <https://doi.org/10.1109/iccsc62074.2024.10617404>
- [2] Fanani, A.Z. and Syarif, A.M. (2023) Historical Building 3D Reconstruction for a Virtual Reality-Based Documentation. *International Journal of Advanced Computer Science and Applications*, **14**, 811-818. <https://doi.org/10.14569/ijacsa.2023.0140985>
- [3] Pérez Nava, F., Sánchez Berriel, I., Pérez Morera, J., Martín Dorta, N., Meier, C. and Hernández Rodríguez, J. (2023) From Maps to 3D Models: Reconstructing the Urban Landscape of San Cristóbal De La Laguna in the 16th Century. *Applied Sciences*, **13**, Article No. 4293. <https://doi.org/10.3390/app13074293>
- [4] Nakao, M. (2023) Medical Image Synthesis and Statistical Reconstruction Methods. *Advanced Biomedical Engineering*, **12**, 21-27. <https://doi.org/10.14326/abe.12.21>
- [5] Jun, W., Son, M., Yoo, J. and Lee, S. (2023) Optimal Configuration of Multi-Task Learning for Autonomous Driving. *Sensors*, **23**, Article No. 9729. <https://doi.org/10.3390/s23249729>
- [6] Clotet, E. and Palacín, J. (2023) SLAMICP Library: Accelerating Obstacle Detection in Mobile Robot Navigation via Outlier Monitoring Following ICP Localization. *Sensors*, **23**, Article No. 6841. <https://doi.org/10.3390/s23156841>
- [7] Pütz, S. (2023) Navigation Control and Path Planning for Autonomous Mobile Robots. *KI-Künstliche Intelligenz*, **37**, 183-186. <https://doi.org/10.1007/s13218-024-00836-x>
- [8] Zhou, L., Wu, G., Zuo, Y., Chen, X. and Hu, H. (2024) A Comprehensive Review of Vision-Based 3D Reconstruction Methods. *Sensors*, **24**, Article No. 2314. <https://doi.org/10.3390/s24072314>
- [9] Xu, Q., Mu, T. and Yang, Y. (2023) A Survey of Deep Learning-Based 3D Shape Generation. *Computational Visual Media*, **9**, 407-442. <https://doi.org/10.1007/s41095-022-0321-5>
- [10] Mescheder, L., Oechsle, M., Niemeyer, M., Nowozin, S. and Geiger, A. (2019) Occupancy Networks: Learning 3D Reconstruction in Function Space. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, 15-20 June 2019, 4455-4465. <https://doi.org/10.1109/cvpr.2019.00459>
- [11] Mittal, P., Cheng, Y., Singh, M. and Tulsiani, S. (2022) AutoSDF: Shape Priors for 3D Completion, Reconstruction and Generation. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 18-24 June 2022, 306-315. <https://doi.org/10.1109/cvpr52688.2022.00040>
- [12] Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R. and Ng, R. (2021) NeRF: Representing

- Scenes as Neural Radiance Fields for View Synthesis. *Communications of the ACM*, **65**, 99-106.  
<https://doi.org/10.1145/3503250>
- [13] Fan, H., Su, H. and Guibas, L. (2017) A Point Set Generation Network for 3D Object Reconstruction from a Single Image. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 2463-2471. <https://doi.org/10.1109/cvpr.2017.264>
- [14] Groueix, T., Fisher, M., Kim, V.G., Russell, B.C. and Aubry, M. (2018) A Papier-Mache Approach to Learning 3D Surface Generation. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 216-224. <https://doi.org/10.1109/cvpr.2018.00030>
- [15] 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>
- [16] Gu, A. and Dao, T. (2023) Mamba: Linear-Time Sequence Modeling with Selective State Spaces.
- [17] Zhu, L., Liao, B., Zhang, Q., Wang, X., Liu, W. and Wang, X. (2024) Vision Mamba: Efficient Visual Representation Learning with Bidirectional State Space Model. *Proceedings of Machine Learning Research*, Vol. 235, 62429-62442.
- [18] Yu, Z., Chen, A., Huang, B., Sattler, T. and Geiger, A. (2024) Mip-Splatting: Alias-Free 3D Gaussian Splatting. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 16-22 June 2024, 19447-19456. <https://doi.org/10.1109/cvpr52733.2024.01839>
- [19] Sobel, I., and Feldman, G. (1968) A 3x3 Isotropic Gradient Operator for Image Processing. The Stanford Artificial Intelligence Project (SAIL).
- [20] <https://github.com/vsitzmann/scene-representation-networks>
- [21] Sitzmann, V., Zollhöfer, M. and Wetzstein, G. (2019) Scene Representation Networks: Continuous 3D-Structure-Aware Neural Scene Representations. *33rd Conference on Neural Information Processing Systems (NeurIPS 2019)*, Vancouver, 8-14 December 2019.
- [22] Reizenstein, J., Shapovalov, R., Henzler, P., Sbordone, L., Labatut, P. and Novotny, D. (2021) Common Objects in 3D: Large-Scale Learning and Evaluation of Real-Life 3D Category Reconstruction. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 10-17 October 2021, 10881-10891. <https://doi.org/10.1109/iccv48922.2021.01072>
- [23] <https://github.com/facebookresearch/co3d>
- [24] Szymanowicz, S., Rupperecht, C. and Vedaldi, A. (2024) Splatter Image: Ultra-Fast Single-View 3D Reconstruction. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 16-22 June 2024, 10208-10217. <https://doi.org/10.1109/cvpr52733.2024.00972>
- [25] Yu, A., Ye, V., Tancik, M. and Kanazawa, A. (2021) pixelNeRF: Neural Radiance Fields from One or Few Images. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 20-25 June 2021, 4576-4585. <https://doi.org/10.1109/cvpr46437.2021.00455>
- [26] Lin, K., Yen-Chen, L., Lai, W., Lin, T., Shih, Y. and Ramamoorthi, R. (2023) Vision Transformer for NeRF-Based View Synthesis from a Single Input Image. 2023 *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, 2-7 January 2023, 806-815. <https://doi.org/10.1109/wacv56688.2023.00087>
- [27] Guo, P., Bautista, M.A., Colburn, A., Yang, L., Ulbricht, D., Susskind, J.M., *et al.* (2022) Fast and Explicit Neural View Synthesis. 2022 *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, 3-8 January 2022, 11-20. <https://doi.org/10.1109/wacv51458.2022.00009>
- [28] Jang, W. and Agapito, L. (2021) CodeNeRF: Disentangled Neural Radiance Fields for Object Categories. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 10-17 October 2021, 12929-12938. <https://doi.org/10.1109/iccv48922.2021.01271>
- [29] Szymanowicz, S., Rupperecht, C. and Vedaldi, A. (2023) Viewset Diffusion: (0-)Image-Conditioned 3D Generative Models from 2D Data. 2023 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Paris, 1-6 October 2023, 8829-8839. <https://doi.org/10.1109/iccv51070.2023.00814>
- [30] Gu, J.T., Trevithick, A., Lin, K.E. *et al.* (2023) NerfDiff: Single-Image View Synthesis with NeRF-Guided Distillation from 3D-Aware Diffusion. <https://arxiv.org/abs/2302.10109>