DOI：10.13245/j.hust.260130

基于混合三平面高斯溅射的高效单视图车辆重建

时　愈[1]　曾子其[1]　刘　照[2]　黄正华[3,4]

(1．武汉工程大学电气信息学院，湖北 武汉 430074；2．中冶武勘工程技术有限公司，湖北 武汉 430081；3．武汉科技大学计算机科学与技术学院，湖北 武汉 430065；4．智能信息处理与实时工业系统湖北省重点实验室，湖北 武汉 430065)

摘要　为解决现有车辆重建存在的精细纹理丢失和体渲染计算量大等问题，提出一种基于混合三平面高斯溅射的高效单视图车辆重建方法，即GSCar。GSCar是一种用于单视图车辆重建的前馈推理方法，在训练和验证阶段均无需额外视图或三维真实数据。GSCar构建了混合三平面高斯表示实现单视图车辆重建，该特征结合了显式点云与隐式高斯属性，既支持高效的三维重建，也提升了渲染质量。通过三平面高斯编码器将二维图像特征映射到三维空间，GSCar实现了形状与高斯潜码的解耦，从而优化了全局形状拟合与局部纹理细节。此外，GSCar利用车辆的对称性来稳健地推断被遮挡区域。实验结果表明，GSCar在三维车辆重建上表现出色，显著提高了车辆重建的质量和速度。

关键词　单视图；三维车辆重建；高斯溅射；点云；三平面

中图分类号　TP391　　**文献标志码**　A　　**文章编号**　1671-4512(XXXX)XX-0001-08

Efficient single-view vehicle reconstruction via hybrid triplane gaussian splatting

SHI Yu[1]　*ZENG Ziqi*[1]　*LIU Zhao*[2]　*Huang Zhenghua*[3,4]

(1．School of Electrical and Information Engineering，Wuhan Institute of Technology，Wuhan 430074，China；2．WSGRI Engineering & Surveying Incorporation Limited，Wuhan 430081，China；3．School of Computer Science and Technology，Wuhan University of Science and Technology，Wuhan 430065，China；4．Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System，Wuhan 430065，China)

Abstract　To address issues such as the loss of fine texture details and the high computational cost associated with volumetric rendering in existing vehicle reconstruction methods, this paper proposes an efficient single-view vehicle reconstruction via hybrid triplane gaussian splatting, termed GSCar. GSCar is a feed-forward inference method for single-view vehicle reconstruction, without requiring additional views or 3-dimensional (3D) ground truth during both training and validation phases. GSCar constructs a hybrid triplane Gaussian representation to achieve single-view vehicle reconstruction. The representation combines explicit point clouds with implicit Gaussian attributes, enabling both efficient 3D reconstruction and improved rendering quality. By using a triplane Gaussian encoder to map 2-dimensional image features into 3D space, GSCar decouples shape and Gaussian latent codes, thereby optimizing both global shape fitting and local texture details. Additionally, GSCar leverages vehicle symmetry to robustly infer occluded regions. Experimental results show that GSCar performs excellently in 3D vehicle reconstruction, significantly improving the quality and speed of vehicle reconstruction.

Key words　single-view；3D vehicle reconstruction；gaussian splatting；point cloud；triplane

从单张二维图像重建三维物体一直是计算机视　　觉和图形学领域的长期挑战之一。单张图像不能提

收稿日期　2026-03-18.

作者简介　时　愈(1985-)，女，副教授，E-mail：shiyu0125@163.com.

基金项目　国家自然科学基金资助项目(62201406)；中国五矿集团青年科技基金(2026QNJJB19).

供多视图信息，使得从有限的单视图准确重建完整且高质量的形状和纹理变得极具挑战性。

近年来，很多学者开展了从单幅图像重建三维物体方法的研究。基于神经辐射场[1]的像素级神经辐射场[2](PixelNeRF)，基于编码的神经辐射场[3](CodeNeRF)，以及三平面超分辨率神经辐射场[4](TripoSR)和真实世界三维重建模型[5](Real3D)等方法虽然取得了一定进展，但存在体渲染[6]计算量大，训练与渲染资源消耗高，依赖多视图数据集等问题。而一种基于三维高斯溅射的扩散模型[7](DiffusionGS)虽然能快速生成三维高斯点云，却高度依赖大规模混合训练策略。

为克服上述挑战，本研究提出了一种基于高斯溅射的单视图车辆图像的新颖三维重建方法(GSCar)。该方法借助固定方向距离系数(FDDC)解码器从单视图图像快速拟合车辆完整表面点云作为高斯核中心，并结合3DGS快速可微分光栅化器完成高效训练与渲染。GSCar构建前馈式混合三平面高斯特征表示，融合显式点云与隐式三平面高斯特性，兼顾重建精度与推理渲染效率。同时设计空间一致的三平面特征提取器，将二维图像特征映射至目标中心标准三维空间，实现全局形状潜码与局部高斯潜码的解耦，支撑车辆点云拟合、高斯属性生成及新视图渲染。此外，GSCar利用车辆对称性使外观成分不依赖原始视图，并与目标空间对称性对齐，增强遮挡视图生成的鲁棒性。

1　提出的方法

在本章节中，将介绍用于单视图车辆重建的GSCar，如图1所示，GSCar采用三平面高斯编码器，从单视图车辆图像中分离形状潜码和三平面特征表示。形状潜码通过FDDC解码器处理，以计算固定方向距离系数，该系数用于将初始点云拟合到车辆表面的点云上。拟合后的点云在三平面中进行空间采样，以推导相应的高斯潜码，随后将其与嵌入式摄像机的参数信息一起输入到高斯解码器中，以生成高斯属性。拟合的点云及其高斯属性随后可以投射到任何新视图上。

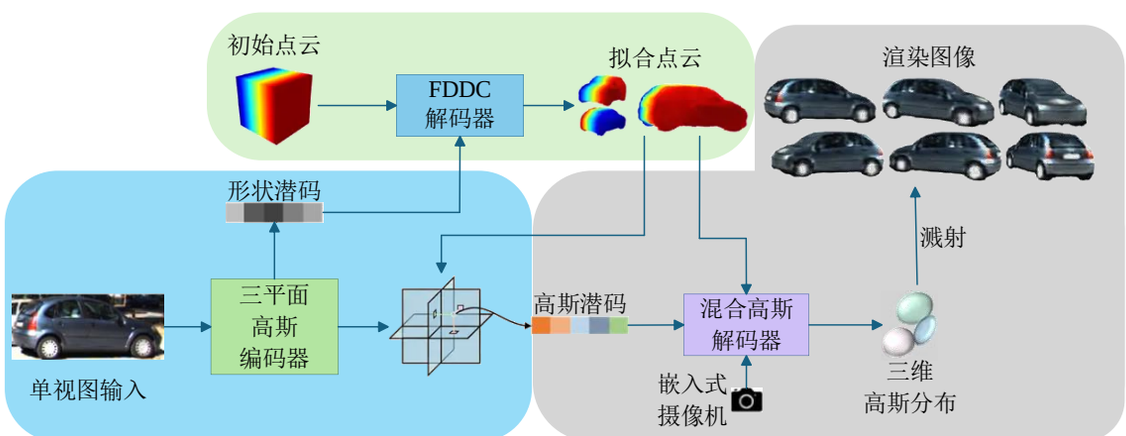


图1　提出的单视图像三维重建的GSCar架构

1.1　预备知识

GSCar的输入包括目标图像$I \in \mathbf{R}^{H\times W\times 3}$以及与其对应的实例分割结果[8]$M \in \mathbf{R}^{H\times W}$和三维对象检测结果[9]$A_{3d}=[x,y,z,l,w,h,r_y] \in \mathbf{R}^7$。这里，$H$和$W$分别表示图像的高度和宽度，$x,y,z$表示目标中心到摄像机视点的三维距离，$l,w,h$表示目标在三维空间中的长度，宽度和高度。$r_y$表示目标相对于$y$轴的旋转角度。

A_{3d}可以在以相机为中心的坐标系内指定目标空间，该空间被转换为以飞行器重心为中心的规范坐标系，主轴对齐于飞行器前进方向，并归一化到[-1,1]范围。在这种配置下，相机在正则坐标系中的位置表示为P_c。此外，定义一个初始点云$P_{init} \in \mathbf{R}^{n\times 3}$，均匀分布于正则空间的表面，其中$n$表示初始化中的点数。点云相对于相机的方向表示为$P_d = P_{init} - P_c \in \mathbf{R}^{n\times 3}$。

1.2　空间一致的混合三平面高斯表示

1.2.1　三平面高斯编码器

在GSCar中，通过三平面高斯编码器对输入的单视图图像信息进行编码，如图2所示。并利用U型网络，将图像特征分解为三平面表示。

该编码器仿照ResNet34架构设计，由多个残差块组成，能够有效提取多尺度图像特征并将其投影到高维表示中。然后，通过多层上采样和卷积层，将编码器提取的多尺度图像特征压缩成包含全局特征的一维形状潜码ϕ_s。ϕ_s被赋予初始点云，并

将其输入到FDDC解码器中，以预测完整的拟合点云P_{fit}，并逐步将编码器提取的特征恢复到原始分辨率。编码器与解码器之间的跳跃连接主要用于传输编码器的底层空间特征，从而通过有意义的结构信息增强三平面特征表示。最终，三平面高斯编码器的输出张量被显式地分为三个不同的平面映射：F_{uv}，F_{uw}和F_{vw}。这些分量被正交化成三平面特征$\phi_{\text{A}} \in \mathbf{R}^{3\times C\times H\times W}$，编码隐式高斯属性。

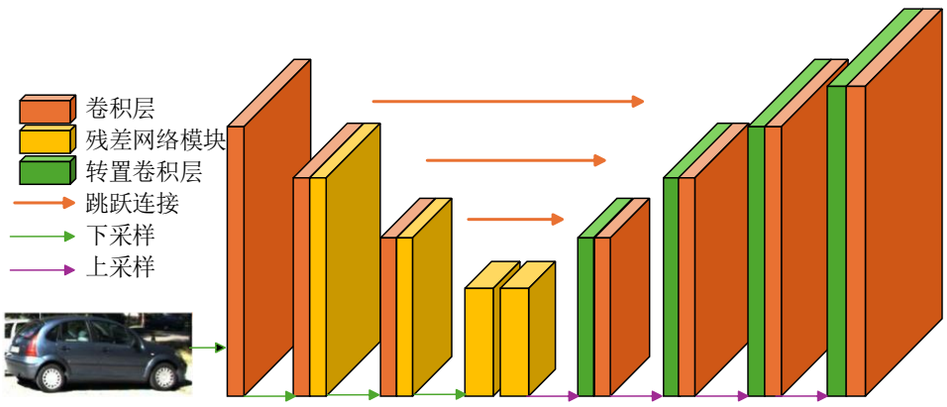


图2　三平面高斯编码器结构

1.2.2　固定方向距离系数(B_{FDDC})

本研究方法作为一种混合表示方法，需要显式的点云作为高斯中心。在这里将介绍如何从单张图像快速拟合完整的车辆点云数据。

首先初始化一个点云$P_{\text{init}} \in \mathbf{R}^{n\times 3}$，其中包含$n$个均匀分布在三维空间表面上的点，其值介于[-1,1]之间，以便与三平面特征信息对齐。为了提升渲染质量并降低计算成本，初始点云的密度配置为长度，宽度和高度的比例为2:1:1。并为每个点引入固定方向距离系数B_{FDDC}，表示相对于原点方向的距离缩放因子。用d和d'表示初始点云和拟合点云相对于原点中心的距离，从而可以推导出$B_{\text{FDDC}} = d'/d$。拟合点云中每个点的坐标由$P_{\text{fit}} = P_{\text{init}} \times B_{\text{FDDC}}$计算。

B_{FDDC}源自解码器Ψ_{FDDC}：

$$B_{\text{FDDC}} = \Psi_{\text{FDDC}}(\phi_{\text{s}}, P_{\text{init}}), \tag{1}$$

初始点云P_{init}的坐标被编码到高维空间中，并与形状潜在变量ϕ_{S}一起，由多层感知机组成的解码器处理，生成每个点的值，并用Sigmoid激活函数将这些值归一化至0到1的范围之间。

此外，由于期望的最终点云是连续闭合曲面，在初始化阶段为每个点云获得N个邻近点。计算完B_{FDDC}后，对邻近点的B_{FDDC}进行平均池化，以确保整个点云在训练迭代过程中趋于平滑。随后计算拟合点云坐标P_{fit}。在渲染过程中，显式点云坐标通过相机外参$\boldsymbol{E}$映射到物体表面。

1.2.3　空间一致性

尽管三平面互相正交，但它们仍然表示固定在原始视图下的二维特征。由于卷积神经网络固有的平移不变性会忽略三维空间相关性，因此在相同的空间坐标上，三平面传递的信息经常不一致，从而最终影响合成图像的精确性。

因此，对于P_{fit}中采样的三平面空间内的每个点，GSCar采用类似于三维卷积的过程以确保空间一致性。鉴于三维卷积本身效率低且占用大量内存，因此通过将点特征聚合替换为在三平面特征内的行和列元素的聚合来改进该过程。这一改进有助于高效并行处理，从而提升计算效率。

更准确地说，对于每个点云$x_i = (u_i, v_i, w_i) \in \mathbf{R}^3$中，通过空间采样操作得到六个向量，分别记为$\boldsymbol{V}_1, \boldsymbol{V}_2, \boldsymbol{V}_3, \boldsymbol{V}_4, \boldsymbol{V}_5, \boldsymbol{V}_6 \in \mathbf{R}^{1\times L\times C}$，$\boldsymbol{V}_1, \boldsymbol{V}_3, \boldsymbol{V}_5$分别表示与三平面特征中点坐标对应的行向量，$\boldsymbol{V}_2, \boldsymbol{V}_4, \boldsymbol{V}_6$则分别表示与三平面特征中点坐标对应的列向量。$i$表示点云的序号，$u_i, v_i$和$w_i$表示对应点云的横、纵和竖坐标，$L$表示行和列的长度，而$C$代表特征通道的数量。点云$x_i$的高斯潜码$\phi_{\text{G}}(x_i) \longleftarrow \phi_{\text{A}}$计算公式为：

$$\phi_{\text{G}}(x_i) = \text{FC}(\text{FC}(\boldsymbol{V}_3 \oplus \boldsymbol{V}_5) \oplus \text{FC}(\boldsymbol{V}_1 \oplus \boldsymbol{V}_6) \oplus \text{FC}(\boldsymbol{V}_2 \oplus \boldsymbol{V}_4)), \tag{2}$$

其中FC表示全连接层,⊕表示拼接函数。这一过程有助于增强各层之间的信息交流，并在保留局部特征和空间关系的同时，改善遮挡区域的成像质量。

在参数量方面，传统的三维卷积网络若要处理具有C个通道的特征体，其单层卷积核的参数量通常为$O(K^3 \times C \times C')$，其中$K$为卷积核大小，$C'$为输出通道数。这种呈立方级增长的参数规模极大地限制了网络的深度和分辨率。相比之下，本研究方法提取三维特征并降维至一维的行和列向量。在公

式(2)中，每个分支首先将对应的两个特征向量沿特征维度展平并拼接，得到一维向量$\boldsymbol{V}_{2LC}$。随后，经过全连接层将其映射至维度D_{h}，三个分支的输出进一步拼接，最终由全连接层映射至输出维度D_{out}。则该模块的总参数量约为$O(3\times\boldsymbol{V}_{2LC}\times D_{\mathrm{h}}+3D_{\mathrm{h}}\times D_{\mathrm{out}})$。由于$L$和设定的$D_{\mathrm{h}}$通常较小，因此参数量比传统的三维卷积小很多。

其次，在FLOPs方面，若直接在宽度、高度和深度分别为W_{g}，H_{g}和Z_{g}的三维空间网格上应用三维卷积，其FLOPs达到$O(W_{\mathrm{g}}\times H_{\mathrm{g}}\times Z_{\mathrm{g}}\times K^3\times C\times C')$，这在处理高分辨率场景时会产生极其庞大的计算量。而对于每个采样的点云x_i，本研究方法仅在其对应的坐标上进行采样和一维向量的正向传播。每个点的计算量主要集中在全连接层的矩阵乘法上，单点的FLOPs约为$O(\boldsymbol{V}_{2LC}\times D_{\mathrm{h}}+D_{\mathrm{h}}\times D_{\mathrm{out}})$，量级远低于传统的三维卷积网络。

1.2.4　混合高斯解码器

混合解码器ψ_{G}通过混合显式的拟合点云P_{fit}与隐式的高斯潜码ϕ_{G}，实现了对车辆属性的深度解析。这种混合表示的方法确保了生成的每个高斯属性既具备精确的三维空间定位，又承载了丰富的局部纹理特征，从而在单视图输入下仍能保持较高的空间一致性。

具体而言，P_{fit}首先利用外参矩阵$\boldsymbol{E}$和A_{3d}计算相对于视点中心的方向P_{d}。随后，P_{fit}和P_{d}被嵌入到高维空间，并与ϕ_{G}一起输入到由多层感知机组成的ψ_{G}中。该过程最终得到高斯属性，包括不透明度$\alpha\in\mathbf{R}^{+}$，协方差矩阵$\boldsymbol{\Sigma}$(由尺度$s\in\mathbf{R}^3$和旋转四元数$q\in\mathbf{R}^4$参数化)以及颜色$c\in\mathbf{R}^3$。整个过程可以表

示为：

$$(\alpha,s,q,c)=\psi_{\mathrm{G}}(\phi_{\mathrm{G}},P_{\mathrm{fit}},P_{\mathrm{d}}),\tag{3}$$

这些高斯参数通过非线性函数进行激活，其中q被归一化，而α,s和c则通过Sigmoid函数被限制在[0,1]的范围内。与使用球面谐波函数表示各向异性颜色信息的原始3DGS不同，GSCar将点云的视图相关的RGB颜色信息编码在高斯解码器本身中。

1.3　对称性增强

GSCar利用混合三平面高斯表示从图像中提取局部特征，有效缓解了现有单视图重建方法过度依赖全局特征的问题。考虑到车辆是具有对称性的物体，这种结构对称性可以有效弥补单视图图像中缺失的三维特征。此外，仅依赖单一视图意味着网络在训练过程中只能使用原始视图进行验证，这可能会阻碍其收敛。当视图无法捕捉到整个场景时，网络倾向于将复杂的三维信息简化为平面表示[10]。因此，在GSCar中，假设车辆作为本质上对称的物体，从左右两侧对应的空间位置观察时应该呈现相似的外观。

由于三平面特征是使用与观察方向正交的图像数据合成的，因此直接在三平面上施加对称性可能会在相机空间中引入无关特征。在这种情况下，卷积神经网络的平移不变性所生成的信息可能与位置和方向嵌入不一致。如果没有来自其他视图的验证，这可能导致网络生成不受控制的表示。

GSCar通过在训练的渲染阶段从两个对称视图生成图像来解决这一挑战，并将这些图像共同纳入损失函数的计算中。由于快速可微分光栅化器的高效性，这一步骤产生的计算开销极小。因此，在目标空间中合成三平面特征时，对称性约束自然地得以应用，从而增强了GSCar用原始视图生成不可见区域视图的鲁棒性。

1.4　损失函数

NeRF方法每次只能渲染有限数量的光线，而GSCar方法可以在每次训练迭代中渲染完整的图像。因此使用均方误差、结构相似性及感知损失来监督训练，渲染RGB损失定义如下：

$$\mathcal{L}_{\mathrm{rgb}}=\mathcal{L}_{\mathrm{MSE}}+\lambda_{\mathrm{s}}\mathcal{L}_{\mathrm{SSIM}}+\lambda_{\mathrm{l}}\mathcal{L}_{\mathrm{LPIPS}},\tag{4}$$

该损失函数由均方误差损失$\mathcal{L}_{\mathrm{MSE}}$，结构相似性损失$\mathcal{L}_{\mathrm{SSIM}}$(及其超参数$\lambda_{\mathrm{s}}$)和感知损失$\mathcal{L}_{\mathrm{LPIPS}}$(及其超参数$\lambda_{\mathrm{l}}$)组成。

由于实际车辆的点云信息通常无法获取，因此利用实例分割结果M来指导B_{FDDC}的计算。具体来说，使用相机外参矩阵$\boldsymbol{E}$将拟合的点云P_{fit}投影到相机平面上，并计算投影点云P_{pro}与M之间的平均绝对误差。点云占用损失$\mathcal{L}_{\mathrm{occ}}$定义如下：

$$\mathcal{L}_{\mathrm{occ}}=\frac{1}{n}\sum_{i=1}^{n}\left|M-P_{\mathrm{pro}}\right|,\tag{5}$$

此外，由于单视图验证的局限性以及高斯核的非结构化特性，许多高斯核的颜色在训练过程中仍然处于无监督状态，这可能导致遮挡区域出现颜色失真。因此引入了一种简单而有效的平均颜色损失$\mathcal{L}_{\mathrm{a}}$，以强制使学习到的高斯颜色保持一致性：

$$\mathcal{L}_{\mathrm{a}}=\mathrm{avg}(c)-\mathrm{avg}(I(M)),\tag{6}$$

其中，avg表示平均值计算操作，$I(M)$则表示对输入图像进行实例分割后得到的车辆目标图像，GSCar的最终损失函数$\mathcal{L}$为：

$$\mathcal{L}=\mathcal{L}_{\mathrm{rgb}}+\lambda_{\mathrm{o}}\mathcal{L}_{\mathrm{occ}}+\lambda_{\mathrm{a}}\mathcal{L}_{\mathrm{a}},\tag{7}$$

其中λ_{o}和λ_{a}是超参数。

2 实验

本研究提出的GSCar模型可以仅凭单视图生成高质量的新视图图像，因此适用于实际应用场景。在单视图车辆重建实验中，选择了KITTI[11]和Nuscenes[12]数据集。经过筛选，得到了6 255张KITTI和30 543张Nuscenes车辆图像，每张图像都附带相应的三维物体检测和实例分割结果。本研究使用Adam优化器，在Nvidia Tesla V100-SXM2 16GB GPU上训练了GSCar模型，共进行了2.5×10^5次迭代，批次大小为16。并将学习率在前1×10^5次迭代中设置为5×10^{-4}，并在第1×10^5 - 2.5×10^5次迭代中降低至5×10^{-5}。为了控制形状、不透明度和外观一致性，超参数λ_s,λ_o和λ_a分别被赋值为1，0.2和2。为了进一步提高训练效率，在第1×10^5次迭代中加入了感知损失$\mathcal{L}_{\text{LPIPS}}$，权重为0.01。点云初始化为$2.5 \times 10^4$个点，这些点从一个长方体中均匀采样，保持长，宽，高之比为2:1:1。

本研究对PixelNeRF[2]、CodeNeRF[3]、AutoRF[1]在单视图车辆重建任务上进行了定量比较。除此之外，还与TripoSR[4]，Real3D[5]和DiffusionGS[7]进行了定性比较。本研究使用多种客观指标来评估从初始视图生成的图像与原始图像之间的差异。这些指标包括峰值信噪比PSNR，结构相似性SSIM，学习感知图像块相似度LPIPS[13]和弗雷歇初始距离FID[14]。此外，为了评估GSCar的计算效率，本研究比较了各种方法的帧率。

2.1 单视图车辆重建

表1中的定量分析结果表明，GSCar在多个图像评估指标上优于其他基线算法。凭借空间一致的混合三平面高斯表示，GSCar能够挖掘出与全局特征互补的高区分度细粒度局部特征，从而在保留三维形状信息的同时丰富外观细节。

GSCar利用车辆对称性的先验知识自然地生成对称结果，然而图像掩码预测的不准确性以及车辆图像固有的不对称性会导致评估指标出现误差。为了更全面地评估GSCar的重建性能，分别在包含对称掩码(Ours w/ mask)和不包含对称掩码(Ours w/o mask)的情况下评估了结果。结果表明，使用对称掩码可以提高本研究方法的精度，证明了其对掩码预测精度较低的物体具有鲁棒性。此外，加粗部分表示本研究方法在该指标上优于其他算法，短横线表示该指标在该方法中不可用。

GSCar完全依赖于来自真实场景的单视图车辆图像数据集。在这种情况下，缺乏额外的视图进行验证给新视图图像的定量分析带来了挑战。为了更全面地评估重建质量，对从不同视角下生成的图像结果进行了定性分析。图3展示了本研究方法与当前最先进的方法进行了定性对比，这些方法均仅使用单视图合成出新视图。

结果表明，PixelNeRF在推理过程中有效地利用了局部特征，从而从原始视图生成了高质量的视图。AutoRF将全局特征与位置编码相结合，从而能够在单视图下生成合理的重建结果。然而，由于全局特征的描述能力有限，它无法捕捉精细的细节。TripoSR和Real3D都依赖于LRM，这提高了生成图像的整体质量。然而，这两种方法都不是专门针对车辆目标设计的，其形状重建结果并不完全自然。DiffusionGS主要针对静态物体或场景生成，未明确包含对移动物体(如行人、车辆)的生成机制，因此对其他视角的车辆重建引入了误差。相比之下，GSCar即使在缺乏车辆类型信息的情况下，也能从单张图像中精确重建正确的车辆形状。此外，还评估了本研究方法的运行效率，并将其与其他基线方法进行了比较，如表2所示。所有报告的测量结果均在单个NVIDIA V100 GPU上进行。PixelNeRF，TripoSR，Real3D和DiffusionGS均使用官方发布的代码实现，而AutoRF使用其非官方版本。

结果表明，基于CNN的方法(PixelNeRF、AutoRF和本研究方法)在图像编码方面明显快于基于LRM的方法(TripoSR和Real3D)。此外，与PixelNeRF和AutoRF不同，后者仅在编码阶段对图像特征进行编码，需要在渲染阶段额外消耗计算资源来生成新视图，导致渲染帧率极低。对于基于扩散模型的DiffusionGS，虽然其最终也输出三维高斯进行渲染，但由于其图像到三维的生成过程需要经历极其耗时的多步去噪迭代网络，导致整体编码时间显著增加，无法满足实时推理的需求。而本研究方法在编码阶段直接对点云的相应高斯属性进行编

码。虽然这确实消耗了一些计算资源，但快速可微分光栅化器允许本研究方法直接将高斯点集铺展到新视图的图像平面上，实现了令人印象深刻的580帧/s，从而有效地满足了实时要求。

2.2 消融研究

在表3和图4中展示了消融研究的结果。首先，消融空间一致性模块(w/o SC)，并使用三平面高斯编码器直接对三维高斯潜在编码进行编码。结果表明，由于缺少平面间的通信，可见区域的成像质量下降，遮挡区域的输出过于简单，通常以单一颜

表1　本研究方法与其他基线方法在Nuscenes和KITTI数据集上的定量比较

方法	Nuscenes				KITTI			
	峰值信噪比	结构相似性	学习感知图像块相似度	弗雷歇初始距离	峰值信噪比	结构相似性	学习感知图像块相似度	弗雷歇初始距离
PixelNeRF								
CodeNeRF	18.25	0.459	0.236	160.60	18.32	0.491	0.223	145.10
AutoRF	18.44	0.462	0.241	146.32	19.01	0.574	0.330	-
Ours w/o mask	18.94	0.491	0.223	145.10	17.88	0.623	0.245	164.83
	21.59	**0.788**	**0.193**	**64.09**	**22.26**	**0.832**	**0.149**	**54.35**
Ours w/ mask	**21.85**	**0.790**	**0.191**	**63.61**	**22.33**	**0.833**	**0.148**	**54.33**



(a)输入　(b)PixelNeRF　(c)AutoRF　(d)TripoSR　(e)Real3D　(f)DiffusionGS　(g)本研究方法

图3　本研究方法与其他基线在KITTI(第1、3、4、6行)和Nuscenes(第2、5行)数据集上的定性比较(第7行是第6行图像中的红色检测框区域的放大内容)

表2　运行时间(帧/s)

方法	编码阶段	渲染阶段
PixelNeRF	330	0.54
AutoRF	184	4.46
TripoSR	1.158	1.128
Real3D	0.615	1.603
DiffusionGS	0.04	0.01
本研究方法	80.556	580.47

色呈现而非丰富的细节，从而对整体性能产生负面影响。

接下来，移除GSCar的对称性增强(w/o symmetry)后，观察到各项评估指标均有显著提升，这表明空间一致的三平面高斯编码器能够从原始视角传递丰富的细节。然而，缺乏对称性先验信息导致在其他视角遮挡区域生成形状和纹理时出现不可预测的影响。为了实现车辆外观的连贯重建，牺牲了原始视角下的性能指标，使GSCar能够利用前馈网络生成更符合人类视觉感知的新视角图像。

最后，移除平均颜色损失函数$\mathcal{L}_a$(w/o avgc)导致评估指标略有下降。这是由于在从原始视角生成图像时，对各向同性高斯点进行颜色监督平均值所带来的益处。然而，更重要的影响来自卷积神经网络的平移不变性的固有局限性；如果没有多视图图像监督，网络在失去整体颜色监督后，往往会在完全未知的区域产生偏向白色的输出。

3　结语

本研究提出了一种基于混合三平面高斯溅射的高效单视图车辆重建方法(GSCar)。该方法仅通过单视图即可实现高质量的三维重建与新视图合成。其核心在于构建了一种空间一致的混合三平面高斯

表3　消融实验：单视图车辆重建

方法	峰值信噪比	结构相似性	学习感知图像块相似度
w/o SC	21.397	0.804 6	0.167 3
w/o symmetry	26.071	0.903 0	0.112 9
w/o avgc	22.066	0.824 4	0.159 0
本研究方法	22.260	0.832 0	0.149 0

表示，通过三平面高斯编码器将二维图像特征映射到以目标为中心的规范三维空间，并利用固定方向距离系数快速拟合车辆表面点云。这种表示方法有效解耦了形状潜码与高斯潜码，在保证全局几何结构准确性的同时，显著提升了局部纹理的重建细节与真实感。实验结果表明，GSCar在重建精度和感知质量上均优于现有的一些方法，并在推理速度和渲染速度上实现了数量级的提升，能够有效满足实时应用的需求。



(a) 输入　(b)w/o SC　(c) w/o symmetry　(d)w/o avgc　(e)本研究方法

图4　本研究方法在KITTI(第1、2行)数据集上的消融分析结果

参 考 文 献

[1] MÜLLER N，SIMONELLI A，PORZI L，et al. AutoRF：learning 3D object radiance fi elds from single view observations[C]// Proc of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans：IEEE，2022：3971-3980.

[2] YU A，YE V，TANCIK M，et al. PixelNeRF：neural radiance fi elds from one or few images[C]// Proc of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Virtual：IEEE，2021：4578-4587.

[3] JANG W B，AGAPITO L. CodeNeRF：disentangled neural radiance fi elds for object categories[C]// Proc of the IEEE/CVF International Conference on Computer Vision. Virtual：IEEE，2021：12949-12958.

[4] TOCHILKIN D，PANKRATZ D，LIU Z X，et al. TripoSR：fast 3D object reconstruction from a single image [EB/OL]. [2026-01-23]. https://github. com/VAST-AI-Research/TripoSR.

[5] JIANG H W，HUANG Q X，PAVLAKOS G. Real3D：towards scaling large reconstruction models with real images [C]// Proc of the IEEE/CVF International Conference on Computer Vision. Honolulu：IEEE，2025：5821-5833.

[6] 陈芬，陈学煌，李旭，等. 面向六自由度视频应用的动态场景重建[J]. 华中科技大学学报(自然科学版)，2025，53(3)：48-55.

[7] CAI Y H，ZHANG H，ZHANG K，et al. Baking gaussian splatting into diffusion denoiser for fast and scalable single-stage image-to-3D generation and reconstruction[C]// Proc of the IEEE/CVF International Conference on Computer Vision. Honolulu：IEEE，2025：25062-25072.

[8] MIAO H，LU F，LIU Z，et al. Robust 2D/3D vehicle parsing in arbitrary camera views for CVIS[C]// Proc of the IEEE/CVF International Conference on Computer Vision. Montreal：IEEE，2021：15611-15620.

[9] 彭刚，刘星宇，刘杰，等. 结合深度梯度与对齐融合策略的目标检测方法[J]. 华中科技大学学报（自然科学版），2025，53(11)：15-20.

[10] 庞立泰，谭鑫. 基于深度估计进行高斯初始化的三维占用预测[J]. 华中科技大学学报（自然科学版），2025，53(12)：178-184.

[11] GEIGER A，LENZ P，URTASUN R. Are we ready for autonomous driving? The kitti vision benchmark suite [C]// Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas：IEEE，2012：3354-3361.

[12] CAESAR H，BANKITI V，LANG A H，et al. Nuscenes：a multimodal dataset for autonomous driving[C]// Proc of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle：IEEE，2020：11621-11631.

[13] ZHANG R，ISOLA P，EFROS A A，et al. The unreasonable effectiveness of deep features as a perceptual metric[C]// Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City：IEEE，2018：586-595.

[14] HEUSEL M，RAMSAUER H，UNTERTHINER T，et al. GANs trained by a two time-scale update rule converge to a local nash equilibrium[C]// Proc of the 31st International Conference on Neural Information Processing Systems. Long Beach：Curran Associates Inc，2017：6626-6637.