

AIGC检测全文详细报告



全文AIGC疑似度：75.51%

正文AIGC疑似度（有效送检内容AIGC率）：84.33%

高危AIGC疑似度：19.69%

中危AIGC疑似度：20.52%

低危AIGC疑似度：35.29%

全文字数：9166

总段落数：116

本科生学年论文

题目：基于扩散模型的2D图片到3D生成的研究

姓 名 杨成

学 号 2022117157

专 业 电子信息工程

指导教师 胡琦瑶

西北大学信息科学与技术学院制

诚信声明

本人郑重声明：本人所呈交的毕业论文（设计），是在导师的指导下独立进行研究所取得的成果。毕业论文（设计）中凡引用他人已经发表或未发表的成果、数据、观点等，均已明确注明出处。除文中已经注明引用的内容外，不包含任何其他个人或集体已经发表或在网上发表的论文。

本段AIGC疑似率：100.0%

特此声明。

论文作者签名：

日 期： 2025年 5 月 17

摘要

扩散模型已经成为2D图像生成建模的主流方式。但扩散模型仍需要大量的数据集，而3D模型的数据采集相对困难，而且生成的模型精度欠佳。因此我通过引入一种新的扩散设置来解决第一个挑战，该设置可以端到端进行训练，仅使用放置的2D图像进行监督；第二个挑战是提出一个图像形成模型，将模型记忆与空间记忆解耦。该扩散模型是可扩展的，训练稳健，并且在样本质量和保真度方面与现有的3D生成建模方法具有竞争力。大多数3D生成研究都侧重于将2D基础模型向上投影到3D空间，要么通过最小化2D分数蒸馏采样(SDS)损失，要么对多视图数据集进行微调。如果没有明确的3D先验，这些方法通常会导致几何异常和多视图不一致性。由于3D数据集中纹理多样性有限，以低质量纹理生成为代价。

为了利用这两种方法的优点，拟提出了双向扩散(BiDiff)，结合3D和2D扩散过程的统一框架，分别保留了3D保真度和2D纹理丰富性。此外，由于简单的组合可能会产生不一致的生成结果，需要进一步用新颖的双向引导来桥接它们，该方法也可以被使用作为基于优化的模型的初始化，将实现高质量、多样化和可扩展的3D生成。

关键字：扩散模型 3D生成 双向扩散 深度学习

Abstract

Diffusion model has become the mainstream way of 2D image generation modeling. However, diffusion models still require a large number of data sets, and the data acquisition of 3D models is relatively difficult, and the generated models are not accurate. So I addressed the first challenge by introducing a new diffusion setup that can be trained end-to-end and supervised only with placed 2D images; The second challenge is to come up with an image formation model that decouples model memory from spatial memory. The diffusion model is scalable, training robust, and competitive with existing 3D generative modeling methods in terms of sample quality and fidelity. Most 3D generation research has focused on projecting 2D base models upward into 3D space, either by minimizing 2D fractional distillation sampling (SDS) losses or by fine-tuning multi-view datasets. Without explicit 3D priors, these methods often result in geometric anomalies and multi-view inconsistencies. Due to the limited texture diversity in 3D data sets, it comes at the cost of low quality texture generation.

In order to take advantage of these two methods, BiDiff is proposed to combine a unified framework for 3D and 2D diffusion processes, preserving 3D fidelity and 2D texture richness, respectively. In addition, since simple combinations can produce inconsistent generation results and need to be further bridged with novel bidirectional guidance, this approach can also be used as the initialization of optimization-based models that will enable high-quality, diverse, and scalable 3D generation.





Keywords: Diffusion model; 3D generation; Bidirectional diffusion;
Deep learning

1绪论

1.1研究背景及意义

近年来，随着人工智能（AI）[1]技术的快速发展，生成式人工智能在图像生成领域取得了显著进展，尤其是在从2D图片生成3D模型[3]这一新兴领域，受到了学术界和工业界的广泛关注。2D图片作为现实世界中最常见的视觉形式，包含了丰富的纹理、色彩和结构信息。因此，通过2D图片生成高质量的3D模型，具有重要的研究价值和广泛的实际应用前景。

传统的3D模型生成方法往往依赖于大量的3D数据，而获取这些数据成本高昂且耗时。相比之下，2D图片的获取相对容易，并且现有的大规模2D图像和文本数据集为研究者提供了大量可利用的先验信息。因此，基于2D图片生成3D模型的技术近年来成为研究热点。这一技术不仅减少了对3D数据的依赖，还加速了3D模型的生成过程，显著提高了生产效率。从2D到3D生成技术通常基于多视角图像重建，需要从多个视角拍摄物体的图像，再通过几何算法重建3D模型。这种方法对数据量和物体形状的复杂性要求较高。随着卷积神经网络（CNN）[4]、生成对抗网络（GAN）[5]和神经辐射场（NeRF）[6]等技术的引入，研究者尝试直接使用2D图片进行3D模型生成。尽管NeRF在生成高保真度的3D表示方面表现出色，但其生成速度较慢，优化时间较长，其他方法也存在一定的缺陷和局限。

为了解决这些问题，现存的主流方法有扩散模型、3D原生、3D基于先验和混合3D。此研究主要以扩散模型为主。扩散模型的核心思想是通过逐步引入噪声来扰乱图像或数据分布，再通过逆向去噪生成目标图像或3D模型。在3D内容生成领域，如何兼顾几何结构的准确性与纹理细节的真实感，一直是有待突破的关键挑战。传统方法往往难以在两者间取得平衡——单纯依赖3D模型生成的内容，虽几何结构稳定

，但纹理细节常显粗糙；而基于 2D 图像训练的模型，虽能产出细腻外观，却难以保证跨视角的 3D 结构一致性。随着影视、游戏、工业设计等领域对高质量 3D 内容的需求激增，需要一种能融合 2D 图像丰富纹理信息与 3D 模型几何约束的生成方案。针对这一点，研究提出双向扩散（BiDiff）方法，旨在构建 2D 与 3D 生成能力的协同体系。充分利用 2D 图像数据的海量规模与纹理细节优势，同时弥补其 3D 结构建模的天然缺陷；结合 3D 模型对空间几何的显式约束，解决单一模态在内容生成中的局限性。通过双向协同机制，让 2D 扩散模型专注于细腻纹理的生成，发挥其在大规模图像训练中积累的外观建模能力；3D 扩散模型则聚焦几何结构的优化，确保生成内容的空间一致性与结构保真度。这种跨模态的优势互补，不仅突破了传统方法在真实感与几何精度上的瓶颈，更通过双向引导实现了生成过程的高效协同，为解决 3D 内容生成中 “纹理 -

结构” 的长期矛盾提供了全新路径。

总体而言，2D到3D生成技术不仅在多个领域展现出了广阔的应用前景，还具有显著的学术研究意义。它可以降低对3D数据的依赖、提高创作效率、推动多模态生成技术的发展，并为虚拟现实和增强现实提供支持。通过对该技术的深入研究和应用探索，未来有望进一步突破现有的技术瓶颈，实现更多的创新和应用。双向扩散技术为2D图片生成高质量3D模型提供了新思路和解决方案。随着扩散模型和生成式AI技术的进一步发展，这一领域有望取得更广泛的应用和技术突破。双向扩散模型

（BiDiff）的提出，进一步推动了2D到3D生成技术的发展，通过有效结合2D扩散模型的纹理生成优势与3D扩散模型的几何保真度，实现了高质量的3D内容生成，不仅显著提升了图像生成的细节表现，还在生成速度和效率方面展现了巨大的潜力。

本段AIGC疑似率：67.46%

图 1-1双向扩散模型

1.2国内外研究现状

随着3D领域技术的突破，3D生成式人工智能（AIGC）的“DALL-E时刻”到来。从DreamFields到DreamFusion，再到Magic3D，3D模型生成技术发展迅速。然而，早期方法通常需要为每个3D模型从头开始优化3D表示，这使得生成过程耗时较长。近年来，Point-E[7]、Shap-E[8]和One-2-3-45[9]等新技术大幅提升了生成速度和准确性，不仅缩短了生成时间，还让用户输入变得更加灵活。

本段AIGC疑似率：57.19%

在开放的3D目标生成方面，图像到3D的方法相比文本到3D的同类方法提供了更高的细粒度控制。然而，在3D内容生成领域，现有方法围绕2D先验驱动、3D优化建模、多视图重建等技术路径持续探索，但在生成效率、几何一致性、数据依赖性等核心问题上仍存在显著瓶颈。

本段AIGC疑似率：59.33%

许多现有方法（如DreamFusion、Magic3D）在生成3D模型时需要从头优化每个模型的3D表示，这个过程通常非常耗时。生成一个3D形状可能需要几十分钟到数小时，限制了这些方法的实际应用场景。例如，Magic3D[11]的生成时间为几十分钟，ProlificDreamer可能需要数小时。生成的3D对象在多个视角下形状不一致。这一问题尤其出现在基于NeRF或其他优化的3D表示方法中，导致生成的3D模型在不同视角下出现不连贯或失真现象。传统的多视角图像重建技术往往依赖大量视角图像进行准确的几何推断。然而，获取如此多的图像在实际操作中具有挑战性，尤其是在稀疏图像条件下，模型的表现通常会显著下降。许多通用的NeRF解决方案在处理稀疏数据集时也表现不佳，难以推广到新场景。部分方法（如基于NeRF的优化技术）依赖于耗时的后处理步骤，以提高生成结果的质量。这些后处理步骤增加了整体计算成本，并使得这些方法难以在实时场景中使用。尽管NeRF等方法能够学习场景的3D表示，但它们对3D先验知识的利用较为有限，这使得生成的3D模型在复杂场景下表现不佳。此外，这些方法通常缺乏对3D形状的全局理解，难以在少量数据的情况下生成高质量的模型。

本段AIGC疑似率：69.81%

DreamFusion和Magic3D等早期工作，依赖于2D先验模型（如CLIP和Stable Diffusion）的广泛知识或生成潜力，通过为每个输入文本或图像从零开始优化3D表示来生成模型。这些方法虽然取得了令人印象深刻的成果，但每个形状的优化过程通常非常耗时，可能需要几十分钟甚至数小时。此外，这些方法常面临“多面”或Janus问题，产生从NeRF或三平面表示继承的过饱和颜色和伪影结果，并在不同的随机种子下生成不同的结果。

One-2-3-45通过利用2D扩散模型的先验进行3D生成，尽管有效，但其多视图图像预测不一致，导致3D重建结果受限。现存的One-2-3-45++[10]克服了这一问题，显著提升了鲁棒性和生成质量。该方法首先通过生成一致的多视角图像，确保各视角之间的一致性；接着利用基于3D扩散的模块，以粗到精的方式预测带纹理的3D网格。

本段AIGC疑似率：58.96%

相比传统的多视角立体重建或基于NeRF的技术，One-2-3-45++无需大量输入图像，能够高效生成一致的多视图3D模型。基于2D扩散网络的多视图生成方法通过一致的多视图条件图像，推动3D重建技术向前迈进。同时，3D扩散网络能够充分利用从3D数据中提取的先验信息，大幅提升生成结果的精确性和细节表现。

1.3 本文的主要内容及章节安排

本文构建了一个深度学习模型，专注于融合2D图像与3D生成技术的优势。所提出的方法基于双向扩散（BiDiff）模型，涵盖三大关键部分：细节保留的2D扩散过程、融合3D特征提取，以及光谱注意力机制下的3D生成优化，显著提升3D生成的空间分辨率和纹理细节。

本文分为五个主要章节，系统地探讨基于双向扩散模型的2D图像到3D生成的应用。首先是绪论，介绍了此领域国际国内研究的背景和动态，并阐述了全文的研究重点和框架构成。第二章深入解释了3D生成方式、基础扩散模型的核心概念，以及这些技术在图像生成中的应用。第三章详细描述了基于双向扩散（BiDiff）模型的2D图像到

3D生成的方法，包括模型的总体架构、关键模块的设计及其功能。第四章通过多角度对比实验和消融实验，展示了提出方法的优越性能和关键设计的有效性。最后一章不仅总结了全文的研究成果，还展望了该领域未来的研究挑战与发展趋势。

本段AIGC疑似率：59.02%

2 相关理论基础

2.1 3D生成方式

正如许多先前的研究所表明的那样，2D纹理和3D几何对于3D对象生成都很重要。然而，将3D结构先验和2D纹理先验结合起来是具有挑战性的：将3D和2D生成模型结合到一个单一的内聚框架中并非易事；在训练和推理中，两个生成模型可能会导致相反的生成方向。

本段AIGC疑似率：58.42%

早期的3D生成方法采用各种3D表示，包括3D体素、点云、网格和隐式函数[11]，用于类别级3D生成。这些方法直接在小规模的3D数据集上训练生成模型，因此，生成的对象可能会错过微小的几何结构或失去多样性。虽然最近有大规模的或高质量的3D数据集，但它们仍然比用于2D图像生成训练的数据集要小得多。借助强大的文本到图像合成模型，通过利用2D生成模型，在没有大规模3D数据集的情况下，出现了3D生成的新范式。一行作品利用来自预训练的文本到图像模型(称为CLIP)或2D扩散生成模型的2D先验来指导底层3D表示的优化。然而，这些模型不能保证跨视图3D一致性，并且每个实例的优化方案都存在高计算成本和过度饱和问题。后来，研究人员使用文本代码或深度图对这些模型进行改进，直接对3D分布进行建模以提高多样性。这些方法减轻了视觉伪影，但仍然不能保证高质量的3D结果。另一项工作是直接从3D数据集学习3D先验。由于扩散模型已成为大多数最新生成模型事实上的网络骨干，它已被适应于使用隐式空间(如点云特征、NeRF参数或SDF空间)来学习3D先验。从3D数据集渲染的合成多视图图像也被用于提供跨视图3D一致性知识，从而在给定的单个图像条件下实现图像。

2.2基础扩散模型

扩散模型[12]作为一类基于概率的生成模型，通过渐进式去噪过程生成高保真数据，近年来在 2D 图像合成领域展现出强大能力（如 Stable Diffusion）。其核心原理是通过前向扩散过程逐步向数据中添加高斯噪声直至纯噪声分布，同时训练反向去噪网络学习从噪声中恢复原始数据的条件概率分布。这一框架天然适合处理具有层次化结构的数据，而将其拓展至 3D 内容生成时，需针对性解决 3D 数据的高维度复杂性（如体素网格的立方级内存需求）与跨视角一致性约束等关键问题。

本段AIGC疑似率：84.43%

在 2D 到 3D 的生成任务中，扩散模型需从单视图或多视图 2D 图像中推断三维结构，生成点云、网格、神经辐射场（NeRF）或体素网格等 3D 表示。具体而言，模型需解析 2D 图像中的深度、透视、遮挡关系和形状先验，将其编码为 3D 空间中的几何与外观特征。例如，DreamFusion通过 2D 扩散模型（如 CLIP 引导的文本 - 图像扩散）生成多视角一致性图像，再通过优化 3D NeRF 辐射场拟合这些图像，实现文本到 3D 的生成；HOLODIFFUSION则提出一种混合显式 - 隐式的 3D 特征网格，仅使用带姿势的 2D 图像监督训练 3D 扩散模型，通过可微分渲染确保跨视角几何一致性，突破了 3D 标注数据的限制。

本段AIGC疑似率：78.46%

扩散模型在3D内容生成中通过适配3D数据表示（如显式体素网格、隐式神经辐射场）与构建跨模态协同机制（2D纹理先验与3D几何约束双向引导），突破了传统方法在几何一致性和数据效率上的瓶颈。具体而言，其通过3D-UNet处理体素特征或优化神经辐射场参数，结合可微分渲染技术将3D结构映射为2D图像以构建光度损失，同时利用2D扩散模型生成多视角纹理细节并注入3D几何先验（如体积正则化），实现纹理真实感与空间结构的统一。尽管面临高维度计算效率、复杂场景鲁棒性等挑战，该方向通过轻量化表示（稀疏体素、神经符号表征）与高效跨模态融合，正推动3D生成技术向实时化、精细化发展，为影视、工业设计等领域提供了从2D输入到3D结构化输出的创新解决方案。

2.3本章小结

本章概述了3D生成方式与基础扩散模型。3D生成需结合2D纹理与3D几何，但面临融合挑战。早期方法受限于数据集规模，新范式则利用2D先验优化3D表示，尽管仍有不足。扩散模型作为强大生成工具，被应用于学习3D先验。基础扩散模型通过逐步去噪生成数据，能从2D输入生成匹配的3D输出，适用于2D到3D任务。本章为后续探讨BiDiff模型奠定基础。

本段AIGC疑似率：70.0%

3基于双向扩散（BiDiff）模型的2D图像到3D生成

3.1双向扩散（BiDiff）模型

鉴于样本来自（未知）数据分布，生成建模的任务是查找参数化模型的最好的近似值。扩散模型是一类基于似然的模型，其中心思想是定义正向扩散（噪声）过程为. 噪声过程将数据样本转换为纯噪声，即i.e., . 然后，模型学习相反的过程，该过程从纯高斯样本开始迭代地将噪声样本转换回数据样本。特别是去噪扩散概率模型（DDPM），它使用高斯分布定义噪声过渡，将

本段AIGC疑似率：87.01%

通过使用 reparameterization 技巧，可以很容易地从这个分布中提取样本：

一个模型同样使用高斯分布定义反向降噪步骤：

其中，是具有学习参数的降噪网络。该序列将噪声调度定义为：

使用带有步骤的线性时间表，必须迭代应用降噪以对目标分布进行采样。但是

，为了训练模型，我们可以直接从以下位置提取样本：如：

本段AIGC疑似率：98.42%

使用网络是很常见的预测 Noise 分量而不是 signal 组件在方程 中;其中具有对边缘分布的分数进行建模的解释 最多为缩放常数.相反，我们使用”-公式“，最近在扩散模型蒸馏和使用扩散对文本条件视频进行建模的背景下进行了探索。这种设计选择背后的原因将在后面显现出来。训练。训练”-Formulation 的包括最小化以下损失：

本段AIGC疑似率：100.0%

鼓励对样本进行降噪 预测干净的样本.采样。一旦去噪网络经过训练，则可以先从纯噪声开始进行采样，即，然后迭代优化使用网络的次数，它以 Target Data Distribution 中的样本结束：

本段AIGC疑似率：100.0%

结合2D和3D先验，使用两种格式的混合组合来表示3D对象。将2D扩散过程的引导添加到3D生成过程中。在二维生成过程中加入了三维扩散过程的指导。

本段AIGC疑似率：58.11%

图3-1方法概述

3.2训练数据

训练数据。我们学习过程的输入是一个由个视频序列组成的数据集(每个视频序列描述了同一对象类别的一个实例(例如，汽车、胡萝卜、泰迪熊)。每个视频包含帧对，每个帧由RGB图像及其对应的摄像机姿态组成，表示为摄像机矩阵。我们的目标是训练一个生成模型，其中是3D物体的形状和外观的表示;此外，我们的目标是仅使用2D训练视频来学习这种分布。

本段AIGC疑似率：85.24%

3D特征网格[19]。作为3D表示，我们选取大小为 N^3 的3D特征体素网格，其中包含 N^3 维的潜在特征向量。给定代表某视频对象的体素网格，我们可以通过渲染函数重构视频的任意帧为。

本段AIGC疑似率：68.11%

HoLODIFFUSION训练找到最优模型参数、，使用初始学习率为(每当总损失达到平台时衰减10倍)的Adam优化器，通过最小化光度损失和自引导光度损失的总和，直到达到收敛。

本段AIGC疑似率：100.0%

在每次训练迭代中，我们从随机选择的训练视频s中随机抽取10个源视图，形成辅助特征的网格，对辅助特征进行降噪，形成，再用去噪。之后，在两次引导过程中对进行降噪和再降噪。为了避免在每次训练迭代中进行两次渲染(一次用于，第二次用于)，我们在每次迭代中随机选择以50-50的概率优化或。photometric loss将去噪体素网格的渲染比较到3个不同的目标视图(不同于源视图)。

除了对模型参数、的训练以外还有渲染功能。从等式和得到的可微渲染函数。(8)和(9)采用发射-吸收(EA)射线推进。首先，已知相机参数，从每个像素的渲染图像发射射线。我们在每条射线上以固定的间隔采样个3D点。对于每个点，我们对相应的体素网格特征进行采样，其中代表三线性插值。然后用MLP将特征解码为，并带参数得到密度和RGB颜色每个3D点的3个。MLP的设计使颜色取决于射线方向，而密度不取决于射线方向，类似于NeRF[18]。最后，EA射线推进渲染图的像素颜色，作为采样颜色的加权组合。权重定义为，其中。

本段AIGC疑似率：63.75%

3.3 3D特征网络

作为 3D 表示我们选择特征体素网格大小含 N^3 维度潜在特征向量。给定体素网格表

示特定视频中的对象，我们可以重建任何帧作为通过渲染函数是函数参数。

本段AIGC疑似率：95.22%

如何为分布构建扩散模型要素网格。人们可能会尝试直接应用过去的方法，但这不起作用，因为我们无法访问 `groundtruth` 特征网格用于培训；相反，这些 3D 模型必须在训练时从可用的 2D 图片中推断出来。

本段AIGC疑似率：96.69%

图 3-2扩散模型

3.4本章小结

本章我们详细介绍了双向扩散（BiDiff）模型的工作原理。该模型基于扩散模型的思想，通过定义正向扩散过程和反向降噪步骤，实现了从纯噪声到数据样本的逐步转换。在正向扩散过程中，数据样本被逐渐添加噪声，直至转换为纯噪声。而反向降噪步骤则学习如何从噪声样本中恢复出原始数据。我们使用了预测干净的样本，并通过最小化损失函数来训练降噪网络。

本段AIGC疑似率：96.5%

其次，我们讨论了训练数据的准备。学习过程的输入是包含同一对象类别的多个实例的视频序列。每个视频都包括多个RGB图像[13]及其相应的摄像机姿势。我们的目标是训练生成模型，其中是3D对象的形状和外观的表示形式。为了实现这一目标，我们仅使用2D培训视频来学习分布。

本段AIGC疑似率：76.85%

最后，我们介绍了3D特征网络的设计。我们选择3D特征体素网格作为3D表示，并给出了如何根据给定的体素网格和摄像机姿势重建任何帧的方法。然而，由于我们无法访问`groundtruth`特征网格用于训练，因此这些3D模型必须在训练时从可用的2D图

片中推断出来。本章我们提供了BiDiff模型在2D图像到3D生成任务中的详细实现方法。

本段AIGC疑似率：70.0%

4 实验结果及分析

4.1 实验设置

我们拟计划进行定量评估，然后是可视化样本用于评估代的质量。数据集和基线。在实验中，使用C03Dv2[17]，这是目前可用的最大的真实物体类别飞行视频数据集。该数据集包含不同对象类别的视频，每个视频围绕对象形成一个完整的圆圈，显示对象的所有侧面。此外，数据集还提供了相机姿势和物体前景蒙版。

本段AIGC疑似率：76.14%

我们在实验中考虑了四个类别、对于每一个类别，并具有C03Dv2注释中定义的最高相机云质量分数，以确保干净的地面真实相机姿势信息。我们将之前的研究pi-GAN[15]、EG3D[16]和GET3D[17]作为比较的基准。Pi-GAN生成由mlp表示的辐射场，并对其进行训练建立一个对抗性目标。与我们的设置类似，他们只使用2D图像监督进行训练。EG3D使用由MLP解码的特征三平面作为底层表示，同时需要图像和相机姿势作为训练过程的输入。GET3D是另一个基于gan的基线，它也需要图像和相机姿势进行训练。我们以C03Dv2的面罩形式提供。由于GET3D在管道中应用了一个可变形的行进四面体步骤，因此由它们生成的样本以纹理网格的形式出现。定量评价。我们报告了Frechet初始距离(FID)和Kernel初始距离(KID)，用于评估我们结果的生成质量，如图4-1为实验所得结果。

本段AIGC疑似率：64.62%

图 4-1生成结果展示

注：总共四类，最右侧为原图片，左边四列为各自对应的3D生成。



4.2与其他方法的对比实验结果

四类 C03Dv2 的 FID 和 KID 将我们的 Holodiffusion 与基线 pi-GAN、EG3D、GET3D、 以及 HoloDiffusion (Nobootstrap)。“VP”列表示方法的渲染是否与 3D 视图一致。

本段AIGC疑似率：100.0%

如表1所示，我们的HoLoDIFFUSION产生了比EG3D和GET3D更好的分数。虽然pi-GAN在某些类别上的得分比我们的好，但我们注意到pi-GAN的3D不可知训练过程无法恢复C03Dv2未对齐形状的正确3D结构。因此，在没有3D视图一致性的情况下，pi-GAN产生的3D神经场基本上模拟了2D图像GAN，定性评价。

本段AIGC疑似率：89.54%

HoLoDIFFUSION产生最吸引人的，一致的和现实的样品在所有。进一步分析pi-GAN与我们的视点一致性。很明显，虽然pi-GAN样本的个别视图看起来很真实，但它们的外观与视图的变化不一致。

本段AIGC疑似率：90.0%

表1 -对比结果

4.3消融实验

为了深入理解HoLoDIFFUSION模型中的各个组件对整体性能的影响，我们进行了一系列消融实验。这些实验主要聚焦于模型的3D表示形式和是否引入双向引导扩散。

本段AIGC疑似率：59.18%

首先，我们测试了改变或去除模型中的某些关键组件（如3D表示形式，双向引导扩散）对生成质量的影响。通过对比这些简化模型与完整模型的FID和KID分数（表

2)，在使用混合网络表示和双向引导扩散的情况下，结果表现远优于消融掉这些模块的实验，我们发现这些模块对于提升生成样本的真实性和一致性至关重要。

本段AIGC疑似率：78.59%

表2消融实验结果

HoLoDIFFUSION 通过混合显隐特征网格与引导扩散模型的联合设计，实现了用 2D 图像监督训练 3D 扩散模型，在生成质量、视图一致性和计算效率间取得突破。消融实验表明，两大创新点分别解决了 3D 表示的瓶颈和训练-测试分布差异问题，为后续 3D 生成模型提供了新范式。

本段AIGC疑似率：65.92%

4.4本章小结

本章详细介绍了HoLoDIFFUSION模型的实验设置、与其他方法的对比实验结果以及消融实验结果。在实验中，我们使用了CO3Dv2数据集，并选择了pi-GAN、EG3D和GET3D作为比较的基准。通过定量评估和可视化样本分析，我们发现HoLoDIFFUSION模型在生成质量、一致性和现实性方面均优于其他基线方法。

本段AIGC疑似率：65.85%

具体而言，HoLoDIFFUSION模型在FID和KID分数上取得了显著的优势，这表明其生成的样本更加真实和一致。同时，通过对比pi-GAN等基线方法的视点一致性，我们发现HoLoDIFFUSION模型在生成具有正确3D结构的样本方面表现出色。

本段AIGC疑似率：59.18%

此外，消融实验的结果也进一步验证了HoLoDIFFUSION模型中各个组件的重要性，并帮助我们确定了最优的训练设置。这些实验结果不仅展示了HoLoDIFFUSION模型的优越性，也为后续的研究提供了有益的参考和启示。

总结与展望

在本文中，我创新性地引入了双向扩散概念，这一概念巧妙地将三维与二维扩散过程融合于一个综合性的框架内。双向扩散不仅充分利用了3D和2D基础模型的稳健先验，还显著提升了对于几何形状与纹理特征的理解与泛化能力。进一步提出了HoLODIFFUSION，这是一个无需条件约束的3D一致生成扩散模型，其独特之处在于仅需依赖姿态图像监督即可完成训练。该模型的核心亮点在于一个可学习的渲染模块，该模块与扩散去噪器协同训练，直接在特征空间层面进行操作。此外，我还引入了预训练的特征编码器，有效地将立方体体积的内存复杂度与最终图像渲染的分辨率进行了分离。即便在有限的姿态图像数据集上，我的方法也能在结果的质量和多样性之间达到理想的平衡。

目前，我的方法仍需在训练阶段获取相机信息。为了解决这一问题，我设想可以联合训练一个视点估计器，以精确定位输入图像。然而，这一设想面临的最大挑战在于如何从头开始训练这一模块，因为输入视图的分布很可能并不均匀。展望未来，面临的下一个显著挑战将是测试条件生成的设置，以及将双向扩散运用到图片生成3D当中。

参考文献

Cao, Yihan, Siyu Li, Yixin Liu, Zhiling Yan, Yutong Dai, Philip S. Yu, and Lichao Sun. "A comprehensive survey of ai-generated content (aigc): A history of generative ai from gan to chatgpt." arXiv preprint arXiv:2303.04226 (2023).

Tucci, Grazia, Valentina Bonora, Alessandro Conti, and Lidia Fiorini. "High-quality 3D models and their use in a cultural heritage conservation

project." The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 42 (2017): 687-693.

Chua, Leon O. "CNN: A vision of complexity." International Journal of Bifurcation and Chaos 7, no. 10 (1997): 2219-2425.

Chen, Zhiqin, Vladimir G. Kim, Matthew Fisher, Noam Aigerman, Hao Zhang, and Siddhartha Chaudhuri. "Decor-gan: 3d shape detailization by conditional refinement." In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 15740-15749. 2021.

本段AIGC疑似率: 76.68%

Mildenhall, Ben, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. "Nerf: Representing scenes as neural radiance fields for view synthesis." Communications of the ACM 65, no. 1 (2021): 99-106.

本段AIGC疑似率: 88.42%

Nichol, Alex, Heewoo Jun, Prafulla Dhariwal, Pamela Mishkin, and Mark Chen. "Point-e: A system for generating 3d point clouds from complex prompts." arXiv preprint arXiv:2212.08751 (2022).

Jun, Heewoo, and Alex Nichol. "Shap-e: Generating conditional 3d implicit functions." arXiv preprint arXiv:2305.02463 (2023).

Liu, Minghua, Chao Xu, Haian Jin, Linghao Chen, Mukund Varma T, Zexiang Xu, and Hao Su. "One-2-3-45: Any single image to 3d mesh in 45 seconds without per-shape optimization." Advances in Neural Information Processing Systems 36 (2024).

本段AIGC疑似率: 55.72%

Lin, Chen-Hsuan, Jun Gao, Luming Tang, Towaki Takikawa, Xiaohui Zeng, Xun Huang, Karsten Kreis, Sanja Fidler, Ming-Yu Liu, and Tsung-Yi Lin. "Magic3d: High-resolution text-to-3d content creation." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 300-309. 2023.

本段AIGC疑似率: 56.08%

Markelj, Primož, Dejan Tomašević, Bostjan Likar, and Franjo Pernuš. "A review of 3D/2D registration methods for image-guided interventions." Medical image analysis 16, no. 3 (2012): 642-661.

Chibane, Julian, Thiemo Alldieck, and Gerard Pons-Moll. "Implicit functions in feature space for 3d shape reconstruction and completion." In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 6970-6981. 2020.

本段AIGC疑似率: 84.83%

Wu, Rundi, et al. "Reconfusion: 3d reconstruction with diffusion priors." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2024.

Kumar, Tarun, and Karun Verma. "A Theory Based on Conversion of RGB image to Gray image." International Journal of Computer Applications 7, no. 2 (2010): 7-10.

Karnewar, Animesh, Niloy J. Mitra, Andrea Vedaldi, and David Novotny. "Holofusion: Towards photo-realistic 3d generative modeling." In Proceedings of the IEEE/CVF International Conference on Computer

Vision, pp. 22976–22985. 2023.

本段AIGC疑似率：57.83%

Chan, Eric R., Marco Monteiro, Petr Kellnhofer, Jiajun Wu, and Gordon Wetzstein. "pi-gan: Periodic implicit generative adversarial networks for 3d-aware image synthesis." In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 5799–5809. 2021.

本段AIGC疑似率：76.61%

Bhattacharai, Ananta R., Matthias Nießner, and Artem Sevastopolsky. "Triplanenet: An encoder for eg3d inversion." In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 3055–3065. 2024.

Gao, Jun, Tianchang Shen, Zian Wang, Wenzheng Chen, Kangxue Yin, Daiqing Li, Or Litany, Zan Gojcic, and Sanja Fidler. "Get3d: A generative model of high quality 3d textured shapes learned from images." Advances In Neural Information Processing Systems 35 (2022): 31841–31854.

Verbin, Dor, et al. "Ref-nerf: Structured view-dependent appearance for neural radiance fields." 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2022.

Ji, Yuzhou, et al. "Fastlgs: Speeding up language embedded gaussians with feature grid mapping." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 39. No. 4. 2025.

功能说明：

AIGC检测可有效识别文本是否部分或全部由AI模型生成，检测结果与论文质量无关、仅表示论文中内容片段存在AI生成可能性的概率。

检测范围:

包括但不限于 ChatGPT-3.5、 ChatGPT-4.0、 Gemini、 Claude、 文心一言、 通义千问、 智谱AI、 豆包、 Kimi、 ol-mini、 ol-preview 、 deepseek

PaperDog论文查重官网_AI论文生成_论文降重_AIGC降重 (<https://www.paperdog.net>)

paperdog, 超级AIGC降重, 一键清除AIGC疑似率痕迹 (<https://www.paperdog.net>)

