



· 文化遗产保护 ·

3DTGC: 基于 3D 重建的文字生成中国画元素

胡琦瑶^{1,2,3}, 王旌宇¹, 李腾飞¹, 彭先霖^{4,5}, 曹瑞⁶, 祝轩¹

1. 西北大学 电子信息学院, 陕西 西安 710127; 2. 新型网络智能信息服务国家地方联合工程研究中心, 陕西 西安 710127; 3. 陕西省高校生成式人工智能与混合现实重点实验室, 陕西 西安 710127;
4. 西北大学 艺术学院, 陕西 西安 710127; 5. 陕西省丝绸之路文化遗产数字化保护与传承协同创新中心, 陕西 西安 710127; 6. 西北大学 计算机学院(软件学院), 陕西 西安 710127

摘要 在无法直接获取多视图信息的情况下, 针对中国画的 3D 数据以及对其进行 3D 重建的研究较少, 因此, 提出了一种名为 3DTGC 的创新方法, 用于文字生成中国画元素的 3D 重建。首先, 文生图模块用来从文字生成中国画元素图像。然后, 多视图生成模块负责多视图合成。通过一系列预处理步骤, 该模块从单幅图像创建 6 个固定视角的图像, 这些图像被用于生成局部光场融合数据集。接下来, 生成的局部光场融合数据集被输入到神经辐射场合成模块中, 用于进一步优化 3D 信息的表达。最后, 在网格生成模块中, 基于前两个步骤的结果创建详细的网格结构。与当前最先进的 3D 重建方法相比, 该框架在可视化效果和技术性能上均表现出显著优势。此外, 3DTGC 能够有效解决其他算法在处理传统中国画数据时常见的多面问题, 从而提升了重建质量和鲁棒性。

关键词 三维重建; 中国传统绘画; 神经辐射场; 扩散模型

中图分类号: TP391.41 **DOI:** 10.16152/j.cnki.xdxbr.2026-02-014

3DTGC: Text generation of Chinese painting elements based on 3D reconstruction

HU Qiyao^{1,2,3}, WANG Jingyu¹, LI Tengfei¹, PENG Xianlin^{4,5},
CAO Rui⁶, ZHU Xuan¹

1. School of Electronic information, Northwest University, Xi'an 710127, China;

2. State-Province Joint Engineering and Research Center of Advanced Networking and Intelligent Information Services, Xi'an 710127, China;

3. Shaanxi Key Laboratory of Higher Education Institution of Generative Artificial Intelligence and Mixed Reality, Xi'an 710127, China;

4. School of Art, Northwest University, Xi'an 710127, China;

5. Shaanxi Silk Road Cultural Heritage Digital Protection and Inheritance Collaborative Innovation Center, Xi'an 710127, China;

6. College of Computer Science, Northwest University, Xi'an 710127, China

Abstract Given the limited availability of multi-view information, research on 3D data and 3D reconstruction

收稿日期: 2025-11-20 修回日期: 2026-01-14

基金项目: 国家重点研发计划项目(2024YFF0907604); 国家自然科学基金(62471390, 62406247); 陕西省教育厅科研计划重点项目(24JS052); 考古探测与文物保护技术教育部重点实验室开放课题(2024KFT03)。

第一作者: 胡琦瑶, 女, 博士, 高级工程师, 从事人工智能与数字媒体艺术交叉研究, huqiyao@nwu.edu.cn。

通信作者: 彭先霖, 男, 博士, 副教授, 从事人工智能与数字媒体艺术交叉研究, pxl@nwu.edu.cn。

of Chinese paintings is scarce. Therefore, this paper proposes an innovative method called 3DTGC for the 3D reconstruction of Chinese painting elements generated from text. First, a text-to-image module generates images of Chinese painting elements from text. Then, a multi-view generation module handles multi-view synthesis. Through a series of preprocessing steps, this module creates six fixed-view images from a single image, which are used to generate a local light field fusion dataset. Next, the generated local light field fusion dataset is input into a neural radiation field synthesis module for further optimization of 3D information representation. Finally, a detailed mesh structure is created in the mesh generation module based on the results of the first two steps. Compared with several state-of-the-art 3D reconstruction methods, this framework demonstrates significant advantages in both visualization and technical performance. Furthermore, 3DTGC effectively addresses the multifaceted problem commonly encountered by other algorithms when processing traditional Chinese painting data, thereby improving reconstruction quality and robustness.

Keywords 3D reconstruction; traditional Chinese painting; neural radiance field; diffusion model

从单一视角重建网格结构是计算机视觉领域的重大挑战。这项任务对智能驾驶系统、机器人技术以及增强现实(AR)和虚拟现实(VR)等环境应用至关重要。此外,非专业人员可以通过它获取3D资源^[1]。

最近的3D重建技术进步主要归功于大规模3D数据集^[24]的出现。这些数据集训练技术在生成速度和输出质量方面显示出巨大潜力,与早期数据集相比,这些现代数据集在规模和质量上都有显著提升。然而,这些方法对多种数据类型的适用性受到数据集固有的类别和注释质量的限制。此外,当前3D和二维数据集的规模之间也存在显著差距。

在将2D图像转换为3D表示的过程中,Dreamfusion使用了得分蒸馏采样来调节扩散模型并优化神经辐射场表示^[5]。尽管取得了这些进展,保持3D一致性仍然存在挑战。为了解决这些问题,许多研究提出了增强模型以进行新视图合成^[5-7]。然而,由于高渲染成本、缺乏专业软件和硬件支持以及优化时间长,实际应用受到限制。

中国画元素的3D重建具有重要意义。在数字文化发展中,这种重建允许将传统中国画元素融入游戏、动画和虚拟现实(VR),促进新艺术作品和商业产品的创作。对于文化遗产保护,许多组织使用3D重建技术重新诠释和保护中国文化遗物和历史遗址。这方面的研究不仅有助于可视化历史画作中描绘的古代物品,还能恢复和保护文化遗物,同时支持使用传统中国元素的艺术创作^[8]。

当前中国画的3D重建方法面临众多困难和挑战^[9]。首先,在数据测试方面,传统中国画具

有复杂的布局结构,直接应用于3D重建过程较为困难。因此,使用包含从中国画中提取的离散元素数据集作为输入至关重要。其次,在重建方法方面,当前单视图重建研究通常使用稀疏合成的新视点。该方法可能导致最终输出缺乏3D连续性、对原始图像的忠实度降低以及生成的3D模型不一致。最后,大多数3D重建工作集中于有形物体或使用现有3D数据集,忽略了对古代中国画和其他艺术形式等抽象实体的技术探索^[10]。

本文介绍了一种创新的古代中国画3D重建框架。该框架通过利用神经辐射场增强传统单视图3D重建方法,实现连续视图3D重建。本研究的创新之处在于将新视图合成与古代中国画结合,允许通过新视图3D重建从艺术图像生成3D模型。本文的贡献如下。

1) 提出了一个文字生成中国画3D重建框架。文生图模块使用FHS-Adapter^[11]从文字生成中国画元素图像。多视角生成模块采用现有扩散模型^[7]从通过预处理获得的单元图像中合成新视点。为了解决稀疏视点重建中的3D不一致问题,本文使用神经辐射场合成模块预测连续视点。网格生成模块利用这些连续视点构建LLFF格式的数据集,然后使用Nerf2mesh进行网格重建,产生高质量视觉效果。

2) 提出了3DTCP数据集,该数据集去除了背景干扰,作为3D重建的输入。该数据集为进一步研究提供了高质量、多样化的中国画元素3D数据。

3) 本文方法在神经辐射场合成模块中使用具有深度图像先验的生成网络来解决中国画数据

集中神经辐射场算法的多面问题。在网格生成模块中,双层多层感知器网络从方向向量中提取镜面和漫反射组件,将光照信息整合到纹理生成中,以实现最佳色彩表现。

3DTGC 框架不仅开创了传统中国画元素 3D 重建的创新方法,还提供了全面的流程,将单图像输入转化为详细的 3D 模型。这促进了文化遗产保护和 VR 中更广泛的应用。

1 相关工作

1.1 单图像到 3D 生成

在传统的 3D 建模方法中,研究人员通常依赖专业建模软件,如 Blender 或 Fusion360,从单视图获取 3D 对象。工作流程复杂且繁琐,依赖于操作员的技术技能。随着在 3D 重建工作中有前景的生成模型如 Diffusion 的出现和应用,人们逐渐将 3D 建模任务交给神经网络。

Dreamfusion 中提出了 SDS 损失的概念^[5],它通过扩散模型合成的新视角不断优化神经辐射场作为最终显示结果。尽管这种方法存在 3D 不一致的问题,其神经辐射场表示需要长时间的渲染优化和训练,但它激发了后续的新视角合成工作^[6-7]等以改善 3D 一致性。在此基础上,许多方法^[12-14]通过各种优化的扩散模型生成新视角,在 3D 重建中取得了突破性的结果。然而,在中国画数据测试中,由于扩散模型生成的视角稀疏,使用稀疏视角进行 3D 重建可能会由于缺乏来自不可见视角的信息而导致 3D 连续性问题。

与扩散模型的新视角合成工作相比,神经辐射场可以有效约束渲染新视角中的 3D 一致性问题。因此,基于稀疏视角,使用神经辐射场与稀疏输入生成连续视角弥补信息的连续性,结果是更合理的最终 3D 模型^[12]。

1.2 单图像到神经辐射场

在许多先前的工作^[12-14]中,为了应对稀疏视点带来的有限训练数据,神经辐射场面临着产生准确 3D 解决方案的挑战。因此,越来越多的人对使用原生 3D 模型产生了兴趣。这些模型旨在将单视图映射到 3D 表示。这些直接的 3D 生成模型通常从真实世界扫描数据或合成 3D 数据集中学习,在很大程度上依赖于大规模 3D 数据集的构建,通过 3D 数据集在生成质量和处理速度上取得了显著进展^[15]。然而,当应用于中国画的

3D 重建时,这些现有的 3D 数据集不包含与中国画元素相关的数据。此外,当前 3D 数据集的规模仍然远小于可用的 2D 数据集数量。虽然在某些对象类型上取得了进展,但这些方法不适用于中国画中多样化的主题^[16]。

另一方面,2D 扩散模型具有强大的生成能力,因此,它们在人工智能生成内容中被广泛使用。这些模型利用来自互联网的大规模文本图像数据集,在学习自然几何形状方面表现出独特优势,并在生成 3D 内容方面显示出巨大潜力。在最近的研究中,Zero-1-to-3^[6]进一步利用潜空间条件扩散模型的微调,并使用合成数据集学习相对摄像机视点的控制。虽然在合成数据集上训练,但该模型可以适应超出分布的中国画。然而,Zero-1-to-3 生成的图像在 3D 重建中缺乏 3D 一致性,基于此,Zero123++^[7]改变了训练策略,在生成图像时考虑联合概率密度,成功增强了 3D 一致性。然而,由于扩散模型的性能有限,它只能生成 6 个新视点^[17]。

传统的神经辐射场算法通常需要几十到几百张图像进行训练,表现出高保真度图像合成,但在稀疏输入视图上面临挑战。在最近的工作中,SparseNeRF 等通过各种策略改善了神经辐射场在稀疏视图输入下的性能,使得可以在稀疏输入视图下获取连续的神经辐射场视图以合成新视角^[18]。

1.3 神经辐射场到 3D 生成

神经辐射场在合成新视图方面取得了显著进展,但在准确表示基础几何形状方面仍面临挑战^[19]。一个主要困难是从神经辐射场中提取网格,这仍然是一个未解决的问题。最近的研究中,NerfMeshing^[20]将训练好的神经辐射场网络整合到 SSAN 模型中,该模型通过估计 TSDF 和外观提取 3D 网格。然而,它并未解决提高神经辐射场网格生成的问题,该问题受到光反射的影响。Make It 3D 引入了一种迭代策略^[20],从神经辐射场渲染中创建干净的点云,将神经辐射场转换为显示表示,避免了噪声点云,但并未专门优化神经辐射场。另一方面,Nerf2mesh 提出了一个两阶段的优化网络,其中第一阶段网络从神经辐射场中提取粗略网格,优化网络进一步增强外观,分别提取漫反射和镜面反射组件进行纹理生成,从而更好地处理光照效果。

在当前相关研究中,Nerf2mesh 利用神经辐射

场优化几何形状、外观和纹理,使中国画中的颜色得以高质量重建。

2 方法

本节将从4个方面介绍3DTGC方法,如图1所示。其中包括:文生图模块;多视角生成模块;神经辐射场合成模块;网格生成模块。

2.1 网络框架

①文生图模块。使用FHS-adapter^[11]进行文字生成中国画元素图像(见图2)。②多视角生成模块。包括从中国画中分割单个元素的模块和生

成多视图图像的模块。③神经辐射场合成模块。使用ZeroRF方法训练神经辐射场,该方法增强了标准神经辐射场技术。神经辐射场是一种基于神经网络的方法,通过从输入图像中学习生成高质量的3D场景表示。它进一步优化了训练,提高了合成复杂场景的效率和准确性。利用这种方法,本文可以从多视图图像中提取3D信息,实现逼真的3D重建。④网格生成模块。包括神经辐射场创建的LLFF格式数据和网格表示生成的部分。LLFF格式数据是通过从多个视点采样场景获得的,这使得3D结构的表示更加准确。

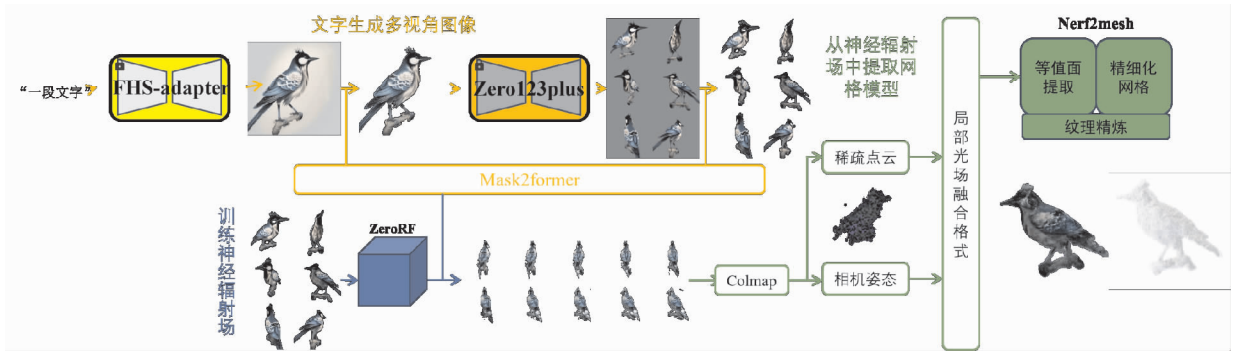


图1 3DTGC的总体架构图

Fig. 1 Overall architecture diagram of 3DTGC

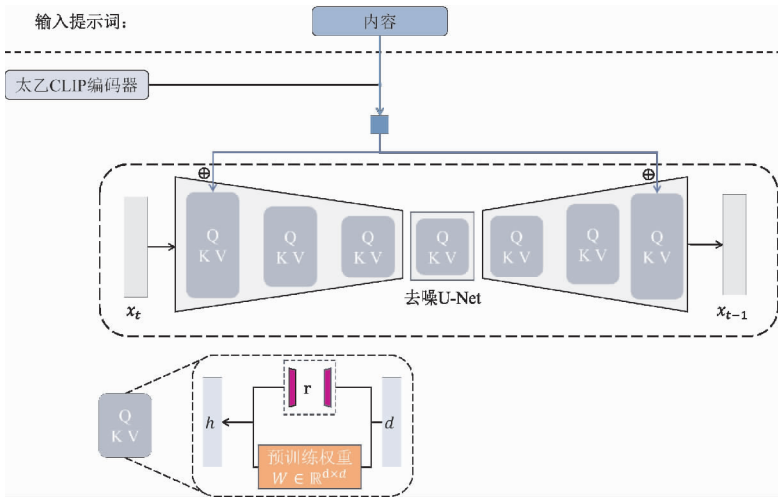


图2 FHS-adapter的总体架构

Fig. 2 Overall architecture of the FHS-adapter

2.2 文生图模块

本文针对中国画独有特征的图像生成任务使用了一种新的细粒度层次语义适配器,即FHS-adapter。由于中国山水画在细节和情境上具有独特的需求,而预训练的扩散模型常常包含大量非山水画数据,使得生成的图像往往缺乏中国山水画的特有风格^[21-22]。为了解决这一问题,FHS-a-

dapter通过批处理方式调整扩散过程,利用外部的细粒度多视角信息来引导图像生成,逐步减少预训练模型中其他风格图像的影响,从而更好地保留山水画元素。

此外,FHS-adapter还替换了原有的编码器,采用专为中文设计的太乙CLIP编码器,以更准确地处理中文文本信息。其使用了一个包含10 000

个高质量图文对的多模态数据集 T2ICLP。

与主流的基于扩散的文到图模型进行比较, FHS-adapter 在模拟中国山水画的笔法(如勾、皴、点、染)、构图空间、元素比例和色彩使用等多个方面表现出色。用户研究表明 FHS-adapter 在重现不同画派和艺术家风格方面具有很高的逼真度。

2.3 多视角生成模块

多视角生成模块的主要功能是生成 6 个多视

图图像,这些图像作为后续模块的关键输入。多视角生成模块每个步骤的输出示例如图 3 所示。对于给定的中国画图像,首先进行几个预处理步骤,以使图像适应多视角生成工作流程的输入要求,并确保与后续神经辐射场处理的兼容性^[23]。这些预处理步骤包括背景干扰减少和分辨率调整,详见图 3 的预处理部分。

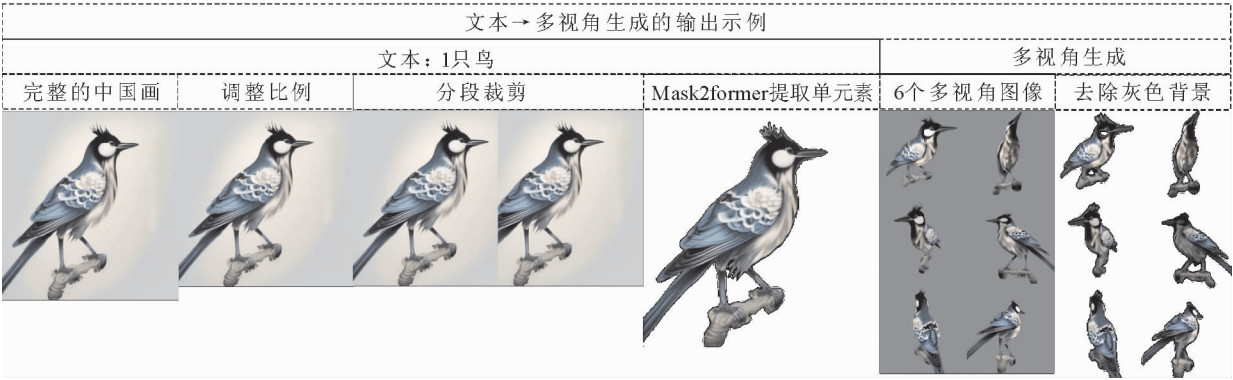


图 3 多视角生成模块流程及各个步骤的输出示例

Fig. 3 Module process and output examples for each step

首先,图像背景可能会通过引入干扰影响重建过程,从而影响后续任务的质量和真实性,例如生成新视图和训练神经辐射场模型。为减轻这种影响,使用诸如 Segment Anything Model (SAM) 等先进工具进行背景分割。

其次,由于数据分辨率不同,预训练的 Zero123 + + 模型无法有效处理高分辨率图像输入。因此,将大图像裁剪以集中目标元素,同时保留重要信息。这一步对于维护 Zero123 + + 的数据输入完整性和确保准确的多视图生成至关重要。

输入的中国画图像被调整为 512 的最近倍数,同时保持原始纵横比。此调整确保与 Zero123 + + 模型的输入要求兼容,并促进后续步骤的高效处理。

如果图像的初始分辨率有任何一边小于 512,背景干扰变得显著,使分割过程复杂化。在此类图像上直接分割和元素提取通常会产生较差的结果。然而,对于数量有限的图像,使用 SAM 或 Photoshop 等工具进行手动裁剪可以绕过分辨率调整步骤,尽管需要增加手动工作量。

对于输入图像 I_{in} 和缩放图像 I_{resize} ,其尺寸为 x 和 y ,使用 $x:y$ 的最简单整数比率 $a:b$ 进行缩放,如方程(1) 所示。这确保了缩放图像的尺寸适合进一步处理。

Resize: $I_{resize}(x, y) \approx I_{in} [512 \times (a), 512 \times (b)]$ (1)

调整纵横比后,图像被裁剪为 512 × 512 分辨率的图块,步长为 256 像素。这种重叠裁剪策略最大化了每个图块中目标元素的存在,提高了 SAM 的分割准确性。本文提供一个示例帮助理解本文的处理方法,如图 3 的多视角生成模块部分所示。假设中国画数据的分辨率为 2 560 × 512。首先,图像分辨率根据近似纵横比 5 : 1 调整为 2 560 × 512,记为 I_{resize} 。然后从左到右以 256 的步长裁剪出大小为 512 × 512 的矩形框,得到 8 个裁剪图块。

使用 SAM^[24] 在透明背景上提取单个元素,利用其高精度分割能力确保干净和准确的提取。SAM 提取的图像随后输入预训练模型 Zero123 + + 进行多视图合成。模型公式如方程(2) 所示:

$I_{out} = M(I_{in})$
 $M \in \{Resize, Divide, SAM, MVG\}$ (2)

其中: M 代表多视角生成处理管道中的操作; I_{out} 和 I_{in} 分别表示最终结果和输入图像。最终生成结果和处理步骤如图 3 的多视角生成模块部分所示。

2.3.1 提取分割图像

为消除背景影响,本文采用 SAM 作为实例分割算法。SAM 是 Meta 发布的图像分割模型,被

广泛应用于众多研究中的各种分割任务。该模型不仅建立了迄今为止最大的分割数据集,还具备优秀的零样本迁移和少样本泛化能力。在面对新的图像分布和任务时,SAM 仍能产生出色的分割结果。由于其独特的零样本迁移能力,SAM 在分割中国画背景时表现出色^[25]。

在上一步中,本文对 n 张图像使用 SAM 进行第二轮精细分割,具体过程包括以下步骤。首先,每张图像输入 SAM 进行实例分割,SAM 生成仅包含单个元素的二进制掩码文件,即背景已移除的部分。为确保分割结果质量,本文设置了过滤条件:分割结果必须占总像素数的至少 10%。该条件有助于排除噪声和小的误分割区域,从而提高分割准确性。在分割过程中,本文将 SAM 的随机种子设置为 64 个均匀分布的种子,此设置有助于确保分割结果的稳定性和一致性,避免因随机因素导致的波动。接下来,本文使用 SAM 生成的二进制掩码文件处理原始图像。具体而言,本文将这些掩码文件应用于原始图像,以获得纯黑背景的单元素图像。通过此步骤,本文成功提取了中国画中的元素,并将它们置于统一的纯黑背景上。因此,完成了中国画元素的初步提取。此过程不仅确保了高质量的分割结果,还为后续的多视图合成和 3D 重建奠定了坚实基础。

2.3.2 生成多视图图像

在使用扩散模型合成 3D 视角的应用中,挑战在于生成的图像缺乏 3D 一致性^[26],这意味着扩散模型为同一物体生成的新视角在细节上有所不同。解决这个问题有两种方法:第一种是使用大规模 3D 数据进行训练,这需要极其复杂的网络和大规模计算资源;另一种方法是 Zero123++ 采用的,通过修改扩散模型的结构,以一对多的方式生成新视角,显著减少训练资源的需求。尽管这种方法相对降低了每个输出的分辨率,但可以显著提高结果的 3D 一致性。在中国画的 3D 重建中,较强的 3D 一致性对于重建的有效性和成功率至关重要。

在中国画中单个元素进行分割图像的下一步中,本文使用 Zero123++ 中的预训练模型合成新视角。3D 一致性的问题也出现在 Zero-1-to-3 中,使得本文无法像神经辐射场那样生成完整的 360° 视角,也阻碍了生成大量用于 3D 重建的多视图。这种方式生成的结果并不理想。

在 Zero-1-to-3 的后续工作中,对 Zero123++

中预训练的扩散模型进行了微调,使其能够一次生成 6 个多视图。它们在保持 Zero-1-to-3 生成能力的同时,在生成过程中考虑了 6 个视图的联合概率密度。扩散模型只能生成 6 个新视角,但仍可以驱动神经辐射场生成更多视角。总之,Zero123++ 可以帮助本文从 6 个固定视角生成具有 3D 一致性的中国画元素的多视图图像,为下一步训练神经辐射场提供输入。

2.4 神经辐射场合成模块

该模块的目的是利用神经辐射场生成扩散模型稀疏视图之间的连续视图。最终结果以 360° 视频的形式呈现。

神经辐射场展示了强大的能力,但容易出现多面问题。本研究认为原因在于神经辐射场模型在面对多样化输入时,单一类型数据集训练的泛化能力较低。ZeroRF 选择固定的深度先验,跳过了数据集训练方法,因此,ZeroRF 在细节上不如在专门数据集上训练的神经辐射场。然而,当处理缺乏训练数据的中国画时,固定深度先验比在其他数据集上训练的先验更不容易遇到多面问题。从视觉角度来看,本研究牺牲了细节生成,但多面问题对网格的损害更不可接受。

ZeroRF 使用预定义的高斯噪声作为深度生成网络的输入,输出 Tensor-Volumetric Model (Tensor-VM) 中的平面和矢量特征。固定的高斯噪声可以引入固定的深度先验,也在后续网络输入中增加随机性。Tensor-VM 是使用张量分解技术表示 3D 场景的框架,有助于高效的数据表示和操作。这些特征用于填充 Tensor-VM 中的网格点。网络为每个网格点生成特定的张量特征,描述 3D 场景中该位置的属性和外观。一旦生成 Tensor-VM,渲染算法可以从任何视点生成 3D 场景的图像。分解的 Tensor-VM 减少了计算和存储需求,同时保持高质量重建。这使得像 ZeroRF 基于神经辐射场的方法在处理大规模场景时更为高效。

对于上一步获得的 6 个多视图图像,本文选择使用 ZeroRF^[27] 中的方法,6 个输入透明背景的多视图图像被训练为神经辐射场表示。神经辐射场渲染的新视角图像均匀分布在 360°,并渲染成一个 100 帧的视频,随后保存为输出结果,这将作为下一步获取网格表示的输入。

2.5 网格生成模块

在此步骤中,本文使用 VT-LLFF 模块获取用于局部光场融合 (Local Light Field Fusion)^[27] 的

数据集格式,然后使用该数据集格式在双阶段网格生成过程中生成最终的网格表示^[28]。网格生成模块的整体处理流程如图 4 所示。

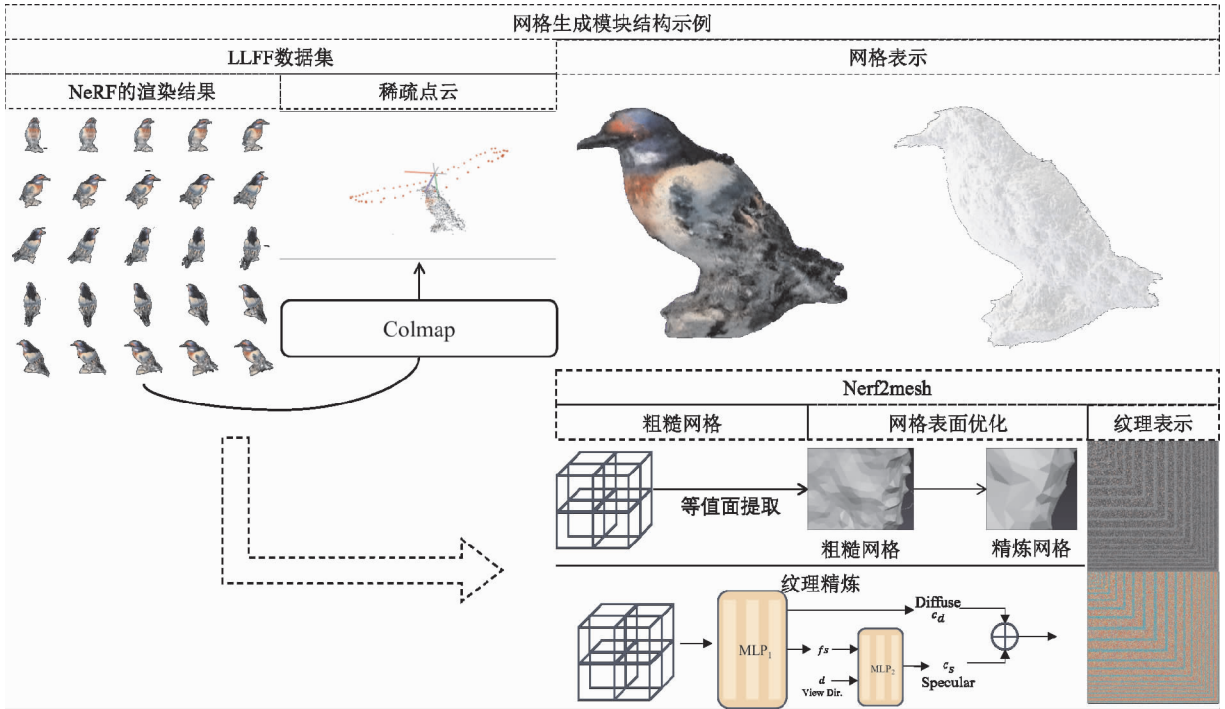


图 4 网格生成模块的总体架构图

Fig. 4 Overall architecture diagram of the mesh generation module

2.5.1 制作 LLFF 数据集

视频到局部光场融合(VT-LLFF)模块的目的是使用神经辐射场创建用于 3D 重建的 LLFF 格式数据集。神经辐射场缺乏明确的 3D 表示,限制了其在实际工程任务中的直接应用。将神经辐射场转换为显式表示仍然具有挑战性,通常涉及生成网格。最近的 3D 重建方法通常在训练期间将神经辐射场参数纳入损失函数,以优化原始网格。该方法增强了模型的细节,但会影响神经辐射场提供的强 3D 一致性,过于关注网格细节可能导致最终输出缺乏整体真实感。

神经辐射场合成模块生成约 360°视角的 100 张图像。VT-LLFF 的目标是利用这些连续渲染的视图创建 LLFF 格式的数据集。最初使用 Colmap 进行特征点提取和匹配,生成稀疏点云。

在特征提取过程中,通常使用尺度不变特征变换(SIFT)算法识别图像中的关键点和描述符。对于每个图像,SIFT 算法通过检测尺度空间中的极值来识别关键点,并通过计算梯度方向直方图生成描述符。设图像为 I ,关键点集为 $\{k_i\}$,描述符集为 $\{d_i\}$ 。特征提取表示为

$$\{(k_i, d_i)\} = \text{SIFT}(I) \quad (3)$$

SIFT 过程包括尺度空间极值检测,其中使用高斯

差分(DoG)在不同尺度上识别关键点。DoG 表达为

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \quad (4)$$

其中: $G(x, y, \sigma)$ 表示高斯核; $I(x, y)$ 是输入图像; $*$ 表示卷积。通过在 DoG 空间中搜索局部极值来识别关键点。随后,通过拟合二次曲面精确定位关键点位置,并消除低对比度和边缘响应点。每个关键点通过计算周围像素的梯度方向直方图分配一个方向,以确定主要方向为

$$\theta(x, y) = \tan^{-1}\left(\frac{L_y(x, y)}{L_x(x, y)}\right) \quad (5)$$

其中: L_x 和 L_y 表示图像梯度。最后,在关键点周围的局部区域中计算梯度方向直方图,以形成描述符。后续的特征匹配过程通过比较不同图像间的描述符来建立对应关系。通常使用最近邻搜索来识别相似的描述符对。设两个图像的描述符集为 $\{d_i\}$ 和 $\{d_j\}$,匹配过程表示为

$$\{(d_i, d_j)\} = \text{Maich}(\{d_i\}, \{d_j\}) \quad (6)$$

特征匹配主要根据描述符相似性在图像间建立对应关系。对于每个描述符,在另一图像中识别最相似的描述符,通常使用欧氏距离或其他相似性度

量。

$$\text{Distance}(d_i, d_j) = |d_i - d_j| \quad (7)$$

同时确保匹配的唯一性和可靠性, d_i 是 d_j 的最近邻, d_j 也是 d_i 的最近邻。特征提取和匹配提供了稳健的几何约束, 有助于在多视图结构光束法(SfM)中进行精确的姿态估计。

在稀疏重建过程中, 使用 SfM 技术估计相机姿态和 3D 结构。利用特征匹配的对应关系, SfM 算法计算相机的外部矩阵(旋转矩阵 $\mathbf{R}$ 和平移向量 $\mathbf{t}$), 以及稀疏 3D 点云。设匹配的特征点对为 $\{(k_i, k_j)\}$, 稀疏重建表示为

$$(\mathbf{R}, \mathbf{t}, \{X_i\}) = \text{SfM}(\{(k_i, k_j)\}) \quad (8)$$

SfM 通过匹配的特征点对恢复相机姿态和 3D 结构。利用基础矩阵或本质矩阵进行相机外部估计, 基础矩阵 $\mathbf{F}$ 满足

$$\mathbf{x}_2^T \mathbf{F} \mathbf{x}_1 = 0 \quad (9)$$

其中: $\mathbf{x}_1$ 和 $\mathbf{x}_2$ 表示两幅图像中匹配特征点的齐次坐标。假设两个相机的投影矩阵为 $\mathbf{P}_1$ 和 $\mathbf{P}_2$, 通过三角测量计算 3D 点 X 为

$$X = \text{Triangulate}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{P}_1, \mathbf{P}_2) \quad (10)$$

通过整合多视图几何关系, 结合所有相机姿态和匹配点生成稀疏点云。随后将这些数据处理为 NumPy 数组, 并计算每个图像的深度范围。通过分析稀疏点云的分布, 结合相机姿态信息进行深度范围计算, 提供更准确的场景尺度表示。相机姿态和深度范围随后创建为 npy 文件。此步骤校正神经辐射场渲染的 3D 形状, 防止因参数过度优化导致的几何失真问题。

VT-LLFF 的输出包括 100 个连续视图图像的稀疏点云和深度范围, 以及相机姿态。这些信息作为 Nerf2mesh 的输入, 通过双阶段网格生成获取网格和纹理。

2.5.2 双阶段网格生成

如何将神经辐射场转换为网格仍然是一个挑战, 本文希望生成的网格能够在真实的 ZeroRF 条件下恢复表面外观, 大多数其他算法生成的纹理只是简单的二维视图, 缺乏真实感。然而, Nerf2mesh^[29] 将纹理分解为镜面反射和漫反射, 漫反射部分模拟光在粗糙表面上的散射, 而镜面反射部分模拟光在光滑表面上的反射。光照条件的加入显著增强了纹理的真实感, 使其更加具有视觉表现力。Nerf2mesh 利用神经辐射场优化几何外观和纹理, 能够高质量地复原中国画的颜色。

Nerf2mesh 的网格结构通常是一个双阶段网

络, 分阶段进行细化。双阶段网格生成的整体框架如图 4 中的 Nerf2mesh 部分所示。在第一阶段, 通过密度网格学习几何形状, 然后初步提取形成粗网格, 在第二阶段进行细化。与输入图像指标相比, 细化后的模型表现出更高的外观和更精细的网格。外观的纹理表示通过两个阶段共享的彩色网格进行学习, 并分解为漫反射和镜面反射成分。收敛后, 导出精细网格, 展开其 UV 坐标, 并将外观烘焙到纹理图像中。

3 实验

3.1 数据集介绍

本文构建了一个专用于 3D 重建任务的高质量中国画元素数据集, 称为 3DTCP。为了确保数据的多样性和代表性, 数据集中的图像主要来源于两个渠道: 一部分是经授权的线下公共博物馆和文化展览馆的高精度扫描资源; 另一部分来自可共享的在线数字艺术数据库。

在数据构建过程中, 本文首先与专业人士合作, 对收集到的画作进行严格的来源鉴定, 确保其具备典型的中国传统艺术特征。随后, 实施了严格的质量控制流程, 剔除了模糊、低分辨率或损坏严重的数据, 形成了初始的高质量图像库。为了适应后续的 3D 重建算法(特别是神经辐射场对清晰几何结构的要求), 本文进一步解决了背景干扰和分辨率归一化的问题, 利用分割算法提取了纯净的画作元素。

最终构建的 3DTCP 数据集总计包含 3 585 幅高分辨率图像。根据题材内容, 本文将其细分为三大类: 包含 361 幅动物题材画作、1 657 幅人物题材画作以及 1 567 幅植物题材画作。该数据集不仅在规模上能够支持深度学习模型的训练, 在内容上也具有极高的多样性, 涵盖了从工笔重彩到水墨写意等多种技法, 以及跨越多个历史朝代和不同流派艺术家的风格。这种丰富的数据分布为验证算法在处理复杂结构和抽象语义时的泛化能力提供了坚实基础。

3.2 对比方法

TripoSR^[14] 是一个基于变压器架构的 3D 重建模型, 能在不到 0.5 s 内从单张图片生成 3D 网格, 并在数据处理、模型设计及训练技术上作出显著改进。

InstantMesh^[11] 是一个从单张图片快速生成

3D 网格的框架,通过结合多视角扩散模型和基于 LRM 架构的稀疏视角重建模型,能在不到 10 s 内创建多样化的 3D 资产,并集成了可微分的等值面提取模块以提高训练效率。

3.3 实验细节

在 CUDA 11.8 环境下,使用单个 NVIDIA RTX 4090 GPU、Ubuntu 20.04 操作系统和 18 核 AMD EPYC 9754 处理器,记录了平均训练时间和最低计算资源需求。结果表明,总共需要 0.5 h,预处理约需 5 min,神经辐射场训练 10 min,网格训练 15 min,至少需要 16 GiB 的 VRAM。由于处理大量点云和图像数据以及复杂的优化,网格训练阶段特别耗时,为提高效率,本文采用低浮点精度训练。本文识别出可扩展性的潜在瓶颈,主要受限于 VRAM 容量和计算能力。随着场景复杂度或分辨率的增加,VRAM 需求显著增长。网格生成的计算复杂度随视图数量线性增加,可能导

致性能问题。因此,建议尽可能使用性能更高的 GPU,并探索分布式计算方法来应对这些挑战。

3.4 定性分析

本文对比较方法进行了视觉分析,如图 5 所示。参考图像作为输入视角的原始视图,关键视觉区域用红色框标出。总体而言,3DTGC 生成的结果更准确地反映了画作在现实中应展现的 3D 形状。与 TripoSR 进行比较时,3DTGC 展现出了显著的优势。尤其在处理不可见视角的三维逻辑性方面,本文的技术能够更加精准地模拟和重建那些从原始视角中无法直接观察到的部分。这意味着生成的三维模型不仅在视觉上更为合理,而且更能够符合物理世界的常规逻辑。此外,与 TripoSR 相比,本文方法在光影效果的处理上也表现得更为出色,能够更加真实地模拟光线如何在不同的表面上反射和折射,从而为用户提供更为生动和真实的视觉体验。

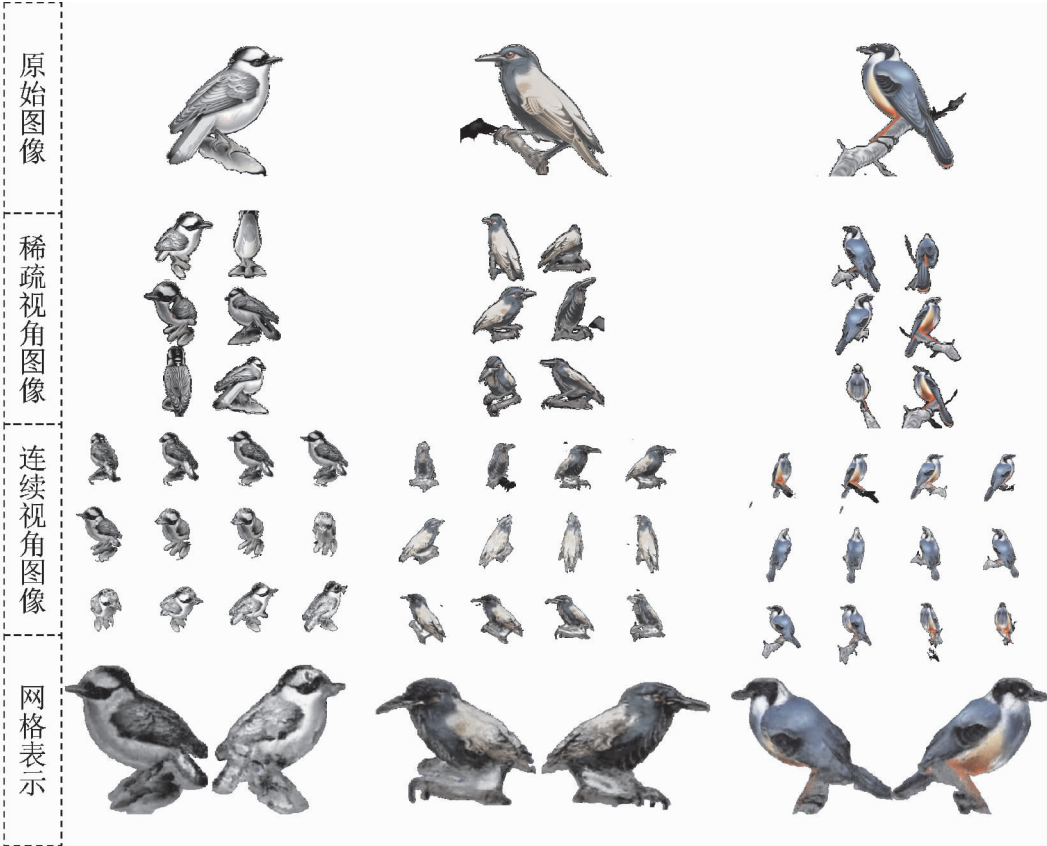


图 5 3DTGC 各个模块输出的可视化结果

Fig. 5 Visualization results of outputs from each module of 3DTGC

与 InstantMesh^[11] 的比较中,3DTGC 同样表现优异。它能够在处理不可见视角时生成更加完整且紧凑的网格结构,显著提高了模型的整体质量和使用的便捷性。更重要的是,通过优化算法和

处理流程,该方法有效避免了在网格生成过程中常见的孔洞和重建失败的问题。Nerf2mesh 的方法在复杂拓扑场景中表现较好,生成高质量的表面网格,纹理图像比其他方法生成的图像更紧凑

和直观。网格的可视化显示在图 6 的第三阶段生成 3D 模型中。在与其他方法^[5,14]的视觉比较中,可以观察到本文的过程生成的 3D 结果具有出色的视觉效果(见图 6)。此外,生成的 3D 形状在 3D 逻辑上更贴近现实世界,生成更美观和自然的 3D 模型。输入 6 个视角后,ZeroRF 的输出显示在图 5 的第二阶段连续视图中。该图展示了 ZeroRF 的渲染结果,包括观音像和鹰的结果。从

图中可以观察到,ZeroRF 的结果表现出优秀的视觉质量和鲁棒性,在大多数情况下,几乎没有不现实的颜色变化,墙壁或漂浮伪影的实例很少或最小。对象的 3D 图像的基本重建与现实世界逻辑一致,没有观察到多面问题。在随后的网格生成模块中,它可以帮助生成具有更强 3D 一致性的网格。

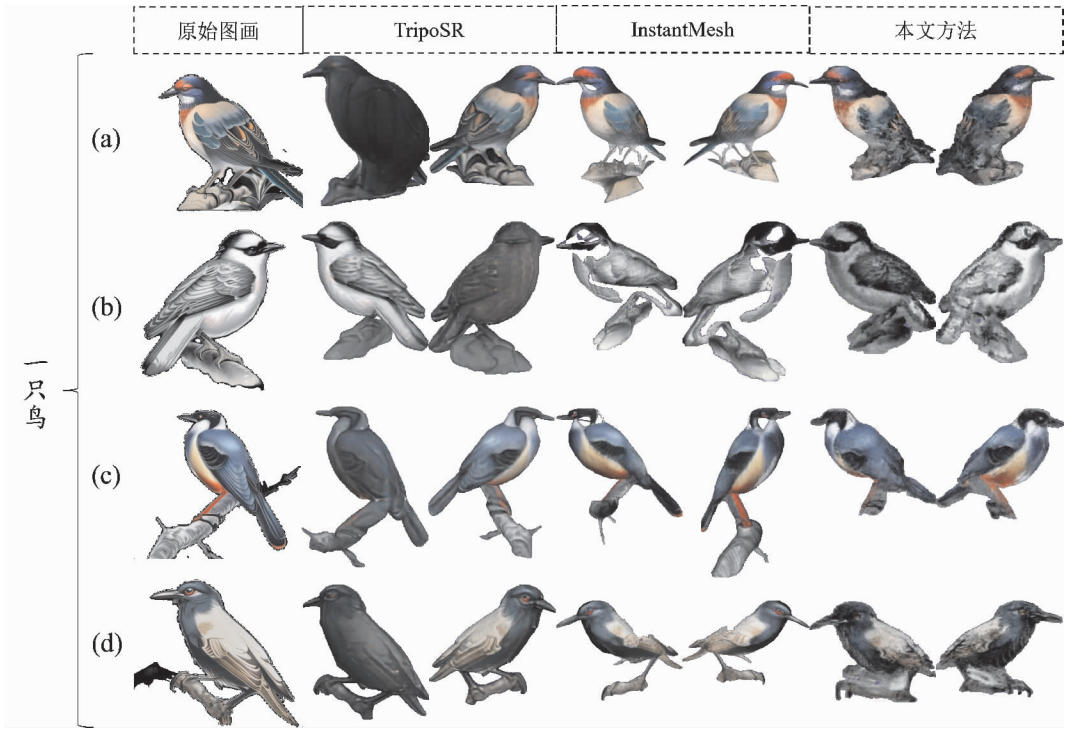


图 6 对比方法进行 3D 重建的可视化结果

Fig. 6 Visualization results of the 3D reconstruction using comparative methods

为了进一步评估不同方法在人类感知层面的表现,本文将提出的 3DTGC 框架与 TripoSR^[14]和 InstantMesh^[5]进行了对比。评估所用的 3D 模型均基于 3DTCP 数据集生成,涵盖了人物、动物和植物 3 个主要类别。

本文招募了 20 名参与者作为本次用户研究的目标群体,为了兼顾技术几何精度与艺术风格忠实度的评估,参与者由两部分组成:10 名计算机视觉领域的专业研究人员和 10 名具有艺术史背景的学生。这种多元化的构成旨在确保评估结果既能反映 3D 重建在结构上的准确性,又能体现对中国传统绘画风格的尊重与还原。

为了确保评估的客观性,本文采用了双盲实验设计,所有生成模型均以随机顺序呈现给参与者,以消除潜在的顺序偏差。本文要求参与者基于以下 3 个关键标准对模型进行评分(见表 1):①几何完整性(评估结构是否完整以及是否存在多面/Janus 问题);②纹理保真度(评估表面细节与输入画作风格的匹配程度);③整体真实感(衡量 3D 元素的视觉合理性)。评分采用 5 点 Likert 量表,其中 1 表示质量较差且存在明显伪影,5 表示高保真度且重建效果逼真。在评分环节开始前,本文提供了明确的书面指导和参考案例,以统一不同背景参与者的评分标准。

表 1 用户研究评分结果(平均值 ± 标准差)
Tab. 1 User study rating results (Mean ± Standard Deviation)

方法	几何完整性	纹理保真度	整体真实感
TripoSR	3. 15 ± 0. 82	3. 28 ± 0. 76	3. 12 ± 0. 79
InstantMesh	3. 62 ± 0. 75	3. 85 ± 0. 68	3. 74 ± 0. 71
本文方法	4. 38 ± 0. 64	4. 52 ± 0. 59	4. 45 ± 0. 61

注:评分基于 1~5 的 Likert 量表,分数越高代表质量越好,最佳结果加粗表示。

用户研究的量化结果如表 1 所示,与定性分析的观察一致,用户评分表明 3DTGC 在重建中国画元素方面优于基准方法。具体而言,本文的模型在纹理保真度方面获得了更高的评分,表明与其他方法相比,本文方法更有效地保留了独特的水墨风格和笔触细节。此外,几何完整性方面的高分反映了本文方法在缓解单视图重建中常见的多面问题上的能力。辐射场合成模块,该算法采用了不同的架构设计。实验结果为了检验方法间差异的显著性,本文对评分数据进行了统计分析,使用配对 t 检验将 3DTGC 的得分分别与 TripoSR 和 InstantMesh 进行比较。结果显示,在所有评估标准上,本文的方法均表现出统计学上的显著优势($p < 0. 05$)。这些统计发现进一步证实了 3DTGC 在生成高质量 3D 中国画元素方面具有卓越性能,能够同时满足技术专家和艺术史学者的审美需求。

3.5 定量分析

如表 2 所示,本文在 4 个对象上进行测试并比较了 3 种最先进的方法。在面数和顶点数上,与 TripoSR 相比,本文方法在图 6(a)上减少了 76. 8%,在图 6(b)上减少了 37. 45%,在图 6(c)上减少了 33. 67%,在图 6(d)上减少了 83. 74%,平均减少了 57. 91%。与 InstantMesh^[11]相比,分别在图 6(a)上减少了 50. 87%,在图 6(b)上减少了 29. 34%,在图 6(c)上减少了 10. 82%,在图 6(d)上减少了 55. 54%,平均减少了 36. 64%。

在 SSIM 和 PSNR 上,与 TripoSR 相比,本文方法在图 6(a)上提升了 0. 049,在图 6(b)上提升了 0. 053,在图 6(c)上提升了 0. 093,在图 6(d)上提升了 0. 095,平均提升了 0. 072 5。与 InstantMesh 比较时,本文方法在图 6(a)上提升了 0. 025,在图 6(b)上提升了 0. 051,在图 6(c)上提升了 0. 084,在图 6(d)上提升了 0. 048,平均提升了 0. 052。

综上所述,在 SSIM 和 PSNR 的性能指标上,

本文的方法不仅在减少模型复杂度和内存使用方面表现优越,同时也在图像质量保真度方面实现了显著的提升,平均在所有图像上相对于 TripoSR 和 InstantMesh^[11]分别提升了 0. 072 5 和 0. 052。这一结果进一步验证了本文方法在保持较低资源消耗的同时,能够有效提高图像的视觉效果。

3.6 消融研究

本文对神经辐射场合成模块进行了消融研究。为了评估其有效性,本文用 OpenLRM 算法采用的版本替换了神经辐射场合成模块,该算法采用了不同的架构设计。实验结果如图 7 所示。OpenLRM^[11]引入了一种在公开的 Objaverse 数据集上训练的架构。然而,由于缺乏专门为中国画策划的 3D 数据,该算法在准确建模这种艺术形式的独特形态和语义方面面临挑战。从新视角看,生成的 3D 模型表现出显著的扭曲,这些输出显示出与感知合理的 3D 几何形态的巨大差异,这表明仅仅在公共数据集上训练神经辐射场合成模块会对最终输出的质量产生负面影响。相比之下,本文的方法使用具有固定深度先验的神经辐射场合成模块,即使在数据稀缺条件下也能在中国画数据上提供改进的泛化。在几何一致性和多面问题的缓解方面,本文的方法产生了最有效的生成结果。

本文在图 8 中对网格生成模块进行了消融研究。图 8 显示了消融后网格下 3D 模型的整体视图,当平滑约束设置为 0 时。在细节部分,“后”表示平滑约束设置为 0 之前的模型细节图像,“前”表示平滑约束设置为 0 之后的模型细节。实验结果表明,当将所有平滑规则设置为 0 时,虽然模型的渲染质量有所提高,但导致生成的 3D 模型表面出现细小的三角形突起,这些突起导致生成模型表面出现孔洞。与消融后的模型相比,原始结果生成更平滑,也能实现更好的视觉质量。

表 2 对比方法的定量指标测量结果

Tab. 2 Quantitative metric measurement results of the comparative methods

方法	对象	面数量	顶点数量	SSIM	PSNR
TripoSR	图 6(a)	101 816	51 005	0. 583	14. 16
	图 6(b)	77 420	38 781	0. 566	13. 63
	图 6(c)	83 958	31 268	0. 523	13. 51
	图 6(d)	101 548	50 860	0. 537	11. 45
InstantMesh	图 6(a)	48 100	24 048	0. 607	13. 87
	图 6(b)	68 588	34 304	0. 568	12. 56
	图 6(c)	62 432	19 094	0. 532	12. 43
	图 6(d)	37 152	18 594	0. 584	12. 22
本文方法	图 6(a)	23 630	11 815	0. 632	14. 44
	图 6(b)	48 458	24 232	0. 619	14. 35
	图 6(c)	55 684	27 800	0. 616	13. 65
	图 6(d)	16 511	8 257	0. 632	13. 78

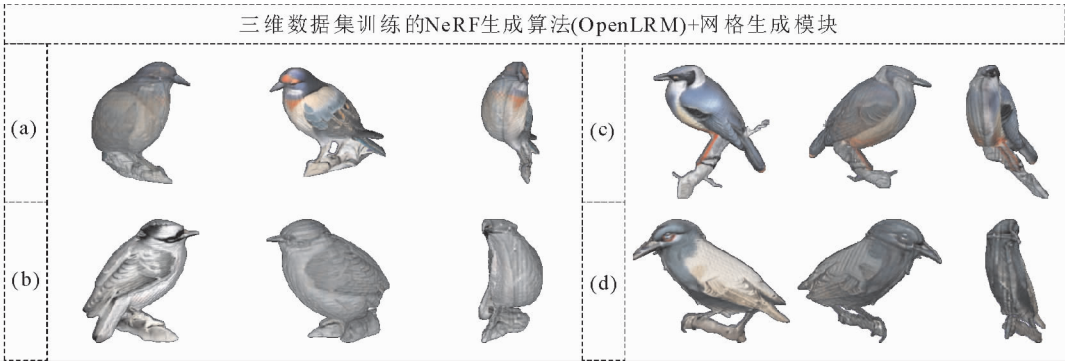


图 7 神经辐射场合成模块消融结果可视化分析

Fig. 7 Visualization analysis of the ablation results of the NeRF synthesis module



图 8 网格生成模型的消融结果可视化分析

Fig. 8 Visualization analysis of the ablation results of the grid generation model

3.7 其他数据集上的实验

本文在壁画数据集上进行了更多实验。实验

结果的可视化分析如图 9 所示。首先,本文的结果可以恢复不同类型壁画上的 3D 形状。此外,

中国画数据集中常见的多面问题在本文的方法中没有出现。本文的结果基本恢复了壁画呈现的 3D 图像,并表现出高度的真实感。壁画结果的测

试指标如表 2 所示。本文保持了与原始数据集相近的指标,并取得了类似的视觉效果,同时仍然保持了相对于基准方法的优势。

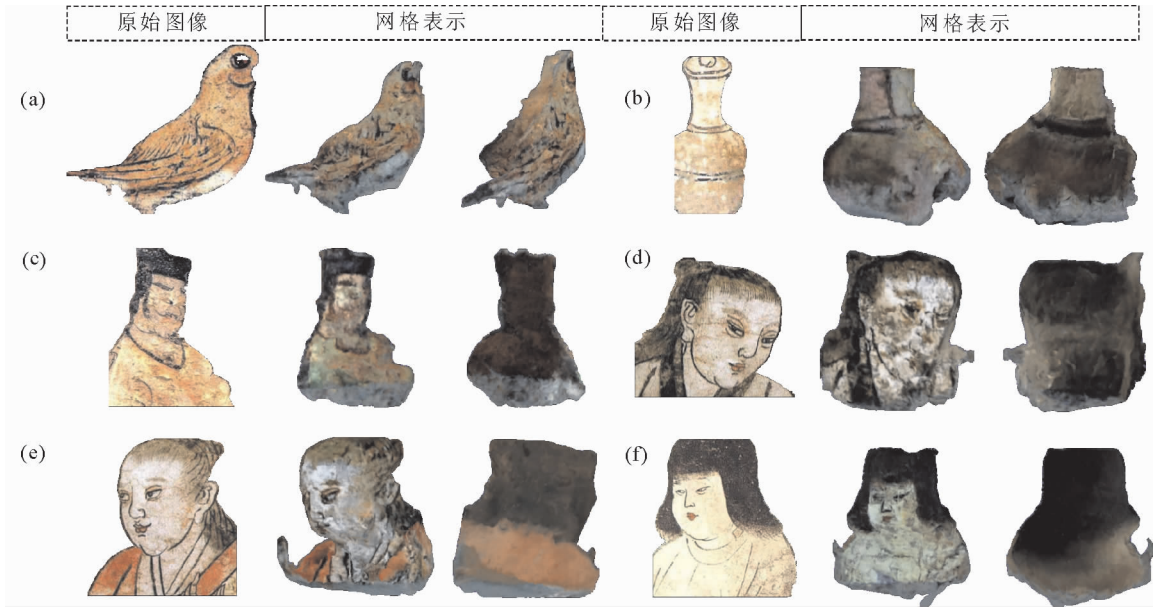


图 9 3DTGC 在壁画数据集上的实验结果可视化分析

Fig. 9 Visualization analysis of the experimental results of 3DTGC on the mural dataset

本文对表 2 中的所有数据进行了显著的统计分析,包括 p 值和置信区间的测试。本文使用其他方法的最小平均值作为假设平均值进行单样本 t 检验。面对各种类型的输入时,本文的方法可以实现整体最佳测试结果,表明本文的方法在统计上是显著的。观察到的结果与零假设有显著差异,表明数据提供了足够的证据来拒绝零假设。

4 结语

本文提出了一种针对中国画元素的 3D 重建框架——3DTGC,利用神经辐射场技术预测连续视角,有效减少传统 3D 重建中的多头和多面问题。为满足算法需求,研究团队设计了专用数据集 3DTCP,并通过实证测试证明了该框架在视觉效果和模型质量上的优势,尤其在深度难以预测的图像中展现出显著的视觉恢复能力。这项技术对文化遗产保护具有重要意义。

尽管 3DTGC 已实现基本的 3D 重建功能,但仍存在局限性。当前模型受限于 Stable Diffusion v1-v2 的架构,图像分辨率不足,未来需探索更先进的生成框架以提升质量。同时,数据集 3DTCP 规模较小且缺乏高质量 3D 数据,限制了模型的

泛化能力。团队计划与相关机构合作,通过高分辨率中国画获取、真实文物 3D 扫描及手工建模扩展数据集,提升模型训练效果。训练过程中还出现过拟合问题,导致噪声斑块和漂浮伪影。为解决这一问题,可采用早停策略和增强正则化技术,进一步提高模型的鲁棒性与泛化能力。

未来研究将聚焦文化遗产保护与游戏产业应用。在文化遗产领域,3DTGC 可用于推断受损文物的原貌,并结合 VR 技术增强博物馆导览体验。在游戏产业中,该框架可协助生成高质量 3D 资产。

参考文献

[1] 徐锦宁,温超,耿国华,等. 面向文化遗产知识传播的百戏俑 VR 体验系统[J]. 西北大学学报(自然科学版), 2025, 55(1): 85-97.
XU J N, WEN C, GENG G H, et al. VR experience system for Baixi terracotta warrior oriented to cultural heritage knowledge dissemination[J]. Journal of North-west University (Natural Science Edition), 2025, 55(1): 85-97.
[2] DEITKE M, SCHWENK D, SALVADOR J, et al. Objaverse: A universe of annotated 3D objects [C] // 2023 IEEE/CVF Conference on Computer Vision and Pattern

- Recognition (CVPR). Vancouver: IEEE, 2023: 13142-13153.
- [3] DEITKE M, LIU R, WALLINGFORD M, et al. Objaverse-xl: A universe of 10m + 3D objects [C] // Advances in Neural Information Processing Systems. New Orleans: NeurIPS, 2023: 35799-35813.
- [4] CHANG A X, FUNKHOUSER T, GUIBAS L, et al. ShapeNet: An information-rich 3D model repository [EB/OL]. (2015-12-09) [2025-06-30]. <https://arxiv.org/abs/1512.03012>.
- [5] POOLE B, JAIN A, BARRON J T, et al. Dreamfusion: Text-to-3D using 2D diffusion [EB/OL]. (2022-09-29) [2025-06-30]. <https://arxiv.org/abs/2209.14988>.
- [6] LIU R, WU R, HOORICK B V, et al. Zero-1-to-3: Zero-shot one image to 3D object [C] // 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris: IEEE, 2023: 9264-9275.
- [7] TAN F, FENG X Y, MA Y P, et al. Two-and three-dimensional deep human detection by generating orthographic top view image from dense point cloud [J]. Journal of Electronic Imaging, 2022, 31(3): 033009.
- [8] 朱耀麟,刘雅琪,杨宇峤. 基于单幅图片的唐代服饰三维重建[J]. 西北大学学报(自然科学版), 2015, 45(3): 379-383.
- ZHU Y L, LIU Y Q, YANG Y Q. 3D reconstruction of Tang Dynasty costume based on single image [J]. Journal of Northwest University (Natural Science Edition), 2015, 45(3): 379-383.
- [9] 王峥嵘,刘鑫达,周明全. 融合扩散模型技术的文物面部三维模型孔洞修补[J]. 西北大学学报(自然科学版), 2025, 55(1): 213-222.
- WANG Z R, LIU X D, ZHOU M Q. Hole repair of 3D cultural relic facial model integrating diffusion model technology [J]. Journal of Northwest University (Natural Science Edition), 2025, 55(1): 213-222.
- [10] 周宇航,冯宏伟,冯筠,等. 基于多视图和显著性分割的古生物三维模型检索[J]. 西北大学学报(自然科学版), 2022, 52(4): 590-601.
- ZHOU Y H, FENG H W, FENG J, et al. 3D ancient organism model retrieval based on multi-view and saliency segmentation [J]. Journal of Northwest University (Natural Science Edition), 2022, 52(4): 590-601.
- [11] XU J L, CHENG W H, GAO Y M, et al. Instant-mesh: Efficient 3D mesh generation from a single image with sparse-view large reconstruction models [EB/OL]. (2024-04-14) [2025-06-30]. <https://arxiv.org/html/2404.07191v2>.
- [12] LIU M, SHI R, CHEN L, et al. One-2-3-45: Any single image to 3D mesh in 45 seconds without per-shape optimization [C] // Advances in Neural Information Processing Systems. New Orleans: NeurIPS, 2023: 22226-22246.
- [13] TANG J, WANG T, ZHANG B, et al. Make-it-3D: High-fidelity 3D creation from a single image with diffusion prior [C] // 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris: IEEE, 2023: 22762-22772.
- [14] TOCHILKIN D, PANKRATZ D, LIU Z, et al. TripoSR: Fast 3D object reconstruction from a single image [EB/OL]. (2024-03-04) [2025-06-30]. <https://arxiv.org/abs/2403.02151>.
- [15] HONG Y C, ZHANG K, GU J X, et al. LRM: Large reconstruction model for single image to 3D [EB/OL]. (2023-11-08) [2025-06-30]. <https://arxiv.org/abs/2311.04400>.
- [16] TANG J X, REN J W, ZHOU H, et al. DreamGaussian: Generative gaussian splatting for efficient 3D content creation [EB/OL]. (2023-09-28) [2025-06-30]. <https://arxiv.org/abs/2309.16653>.
- [17] KEMELMACHER-SHLIZERMAN I. Internet based morphable model [C] // 2013 IEEE International Conference on Computer Vision. Sydney: IEEE, 2013: 3256-3263.
- [18] YANG Y, FENG C, SHEN Y, et al. FoldingNet: Point cloud autoencoder via deep grid deformation [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 206-215.
- [19] WANG N, ZHANG Y, LI Z, et al. Pixel2mesh: 3D mesh model generation via image guided deformation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(10): 3600-3613.
- [20] RAKOTOSAONA M J, MANHARDT F, ARROYO D M, et al. NeRFMeshing: Distilling neural radiance fields into geometrically-accurate 3D meshes [C] // 2024 International Conference on 3D Vision (3DV). Davos: IEEE, 2024: 1156-1165.
- [21] SMITH E, FUJIMOTO S, ROMERO A, et al. Geometrics: Exploiting geometric structure for graph-encoded objects [C] // Proceedings of Machine Learning Research (PMLR). Long Beach: IEEE, 2019: 5866-5876.
- [22] WANG H, DU X, LI J, et al. Score Jacobian chaining: Lifting pretrained 2D diffusion models for 3D generation [C] // 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver: IEEE, 2023: 12619-12629.

[23] GUO S, WANG Q, GAO Y, et al. Depth-guided robust point cloud fusion NeRF for sparse input views [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(9): 8093-8106.

[24] KIRILLOV A, MINTUN E, RAVI N, et al. Segment anything[C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris: IEEE, 2023: 3992-4003.

[25] NIEMEYER M, BARRON J T, MILDENHALL B, et al. RegNeRF: Regularizing neural radiance fields for view synthesis from sparse inputs [C] // 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE, 2022: 5470-5480.

[26] PENG X, HU Q, FAN F, et al. FHS-adapter: fine-grained hierarchical semantic adapter for Chinese landscape paintings generation [J]. Heritage Science, 2024, 12: 265.

[27] MILDENHALL B, SRINIVASAN P P, CAYON R O, et al. Local light field fusion[J]. ACM Transactions on Graphics (TOG), 2019, 38(4): 1-14.

[28] SHI R, WEI X, WANG C, et al. ZeroRF: Fast sparse view 360° reconstruction with zero pretraining [C] // 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2024: 21114-21124.

[29] TANG J X, ZHOU H, CHEN X K, et al. Delicate textured mesh recovery from NeRF via adaptive surface refinement[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Paris: IEEE, 2023: 17739-17749.

(编辑 张 欢)