用于几何与颜色联合修复的掩码条件体素扩散

Aarya Sumuk斯坦福大学机械
工程系 asumuk@stanford.edu

arXiv:2601.00368v1 [cs.CV] 1 Jan 2026

摘要

我们提出了一种轻量级的两阶段框架，用于对受损三维物体进行几何与颜色联合修复，其灵感源于文化遗产文物的数字修复。该流程将损伤定位与重建分开。第一阶段，二维卷积网络对从体素化对象中提取的 RGB 切片预测损伤掩码，并将这些预测汇总为体积掩码。第二阶段，基于扩散的三维 U-Net 直接在体素网格上执行掩码条件修复，在重建几何与颜色的同时保留已观测区域。模型通过复合目标函数联合预测占据情况与颜色，该目标函数结合了占用重建、掩码颜色重建和感知正则化。我们使用标准几何与颜色指标，在一组经过筛选且带有合成生成损伤的带纹理文物上评估该方法。与基于对称性的基线方法相比，在固定的 32^3 分辨率下，我们的方法能够生成更完整的几何结构和更连贯的颜色重建。总体而言，结果表明，显式掩码条件控制是一种实用的方法，可用于引导体积扩散模型进行三维几何与颜色联合修复。

1 引言

重建不完整或受损的三维物体是计算机视觉和图形学领域由来已久的问题，其应用涵盖文化遗产保护、机器人学以及一般的形状理解。在遗产文物的语境下，雕塑、陶瓷器和小型物体的三维扫描通常会因侵蚀、破损或采集过程的局限而包含缺失区域。对这类扫描结果进行数字修复，可以在不接触脆弱物体的情况下实现无损研究、可视化和存档，同时避免可能引入不可逆变化的实体干预。
从技术角度来看，这一问题可以表述为：根据部分观测结果，同时补全三维几何与外观。经典重建方法主要关注几何，而许多现代扫描结果包含颜色或纹理信息，可为修复提供额外背景。与此同时，在高分辨率网格或隐式表示上运行的方法通常需要大量计算资源，这可能限制其在轻量级或探索性场景中的实用性。

本文研究基于体素的表示中的几何与颜色联合修复。我们从受损的带纹理网格出发，将物体体素化为二值占用网格，并生成与之对齐的逐体素 RGB 体积。通过使用逐切片孔洞掩码和形态学腐蚀移除体积中具有结构的部分，合成缺失区域，从而得到成对的受损与完整样本。目标是重建完整的占用网格及其对应的颜色体积，使其尽可能近似原始未受损的形状和外观。

预印本。

为解决这一问题，我们采用两阶段流程，将损伤定位与内容重建解耦。第一阶段中，二维卷积 U-Net 对从体素网格提取的各个 RGB 切片预测损伤掩码。随后汇总切片级预测，生成用于识别需要修复区域的体积掩码。这一步显式定位简化了后续重建任务，使模型能够专注于填补缺失区域，而非检测这些区域。

在第二阶段，我们使用直接作用于体素网格的基于扩散的 U-Net 执行三维修复。该网络以掩码占据、预测的损伤掩码和掩码颜色通道作为输入，并经过训练以重建缺失几何和颜色，同时保留完整区域。我们采用扩散时间步和正弦时间嵌入来引导去噪过程，遵循近期关于用于三维形状生成与补全的扩散模型的研究 Nam 等人[2022], Schröppel 等人 [2024]。训练目标将用于占用重建的二元交叉熵与用于颜色的掩码 L_1 损失相结合，并加入可选的在二维切片上计算的感知正则化，以及针对类陶瓷材料设计的弱颜色先验。

我们在一个小型精选带纹理文物数据集上评估所提出的方法，该数据集由公开可用的三维扫描和 CAD 模型组成。性能采用标准几何指标进行衡量，包括倒角距离和 1 mm 阈值下的 F-score，以及用于颜色保真度评估的掩码均方误差和峰值信噪比。我们将其与简单基线方法进行比较，例如基于对称性的修复，以及先进行泊松曲面重建再进行体素化的方法 Kazhdan 等人 [2006],Kazhdan 和 Hoppe [2013]。在留出样例上，基于扩散的模型相比这些基线能够生成更完整的几何结构和更连贯的颜色重建，同时仍保持计算轻量。

尽管本研究的范围仅限于中等体素分辨率和合成生成的损伤，但结果表明，在受限条件下，显式地以预测的损伤掩码为 3D 扩散模型提供条件，是实现几何与颜色联合修复的可行策略。

2 相关工作

我们的工作与几何重建、体积形状补全、基于扩散的生成建模和外观修复方面的既有研究相关。我们简要回顾这些领域，并说明我们的方法如何融入现有文献。

2.1 几何重建与补全

修复不完整 3D 几何的经典方法依赖表面插值和隐式重建。泊松曲面重建及其屏蔽变体将表面补全表述为对定义在定向点云上的泊松方程求解，从而根据局部数据生成水密网格 Kazhdan et al. [2006], Kazhdan and Hoppe [2013]。这些方法虽然对平滑表面效果良好，但往往会过度平滑高频细节，也不会显式推断缺失区域的位置或结构。

基于学习的方法通过体数据或点云表示来解决形状补全问题。早期的基于体素的模型使用编码器—解码器架构，根据一个或多个视图重建形状 Choy 等人 [2016], ，后续工作则将这些思路拓展至场景尺度补全任务 Dai 等人 [2018]。点云方法直接从无序点集中推断缺失几何 Yuan 等人 [2018], Yang 等人 [2018], ，但通常还需要额外的表面重建步骤才能得到水密网格。

2.2 基于切片的方法与 2.5D 处理

通过二维切片处理体数据已广泛应用于医学成像，在计算效率与空间上下文之间实现了折中。在所谓的 2.5D 方法中，来自多个正交视图的独立二维预测会被汇总，以形成三维结果；这种方法无需承担完整三维卷积的计算成本，也能取得有竞争力的性能 Çiçek 等人 [2016]。

我们的第一阶段采用了类似策略：在从体素化对象中提取的二维RGB切片上预测损伤掩码，并将其聚合为体积掩码。这种对

缺失区域的显式定位简化了后续的三维重建，并使扩散模型能够专注于修复，而非检测。

2.3 用于三维形状补全的扩散模型

近年来，扩散模型已成为一类强大的三维数据生成模型。多项研究在从隐式形状表示中学习到的潜在空间进行扩散，以显著增加的训练复杂度为代价，实现高分辨率形状合成 Nam 等人 [2022]。其他方法则直接对体素网格或占据场应用扩散，在 CAD 风格数据集上取得了出色的补全性能 Schröppel 等人 [2024]。
与这些方法不同，我们的方法在相对较低的体素分辨率下应用扩散，并显式地以预测的损伤掩码为条件。这种设计优先考虑计算简便性和可控性，而非高分辨率合成。

2.4 颜色与外观修复

与几何补全相比，三维数据的外观修复受到的关注相对较少。先前的研究探索了使用感知损失预测网格的颜色或纹理，通常将几何视为固定，并侧重于外观迁移，而非联合重建 Johnson 等人 [2016]。

我们的方法在统一的基于体素的扩散框架中进行几何与颜色联合修复。颜色被预测为以掩码几何和完整外观为条件的残差，并通过掩码重建损失以及在二维切片上使用预训练卷积网络计算的感知特征进行正则化 Simonyan 和 Zisserman [2014]。

3 方法

我们提出一种两阶段流程，使用基于体素的表示，对受损三维物体进行几何与颜色联合修复。所有实验均在固定的空间分辨率 $32 \times 32 \times 32$ 下进行，以实现高效训练和评估。第一阶段通过逐切片二维分割估计体积损伤掩码，第二阶段则使用基于扩散的 U-Net 执行掩码条件三维修复。

3.1 体素表示

每件文物均表示为一个二值占用网格

$$V^{\mathrm{dam}} \in \{0,1\}^{32\times32\times32},$$

并配有与之对齐的 RGB 颜色体积

$$C^{\mathrm{dam}} \in [0,1]^{3\times32\times32\times32}.$$

将网格归一化至单位立方体后进行体素化。通过纹理渲染为占据体素赋予颜色值，而空体素则赋予零颜色。
通过使用逐切片孔洞掩码和形态学腐蚀，从体素网格中移除结构化区域来引入合成损伤。此过程生成成对的受损体积数据（$V_{\text{受损}} C^{\mathrm{dam}}$）和对应的真实完整体积数据（$V_{\text{真实值}} C^{\mathrm{gt}}$），以及一个二值损伤掩码

$$M = \mathbb{1}[V^{\mathrm{gt}} = 1 \wedge V^{\mathrm{dam}} = 0].$$

3.2 阶段一：损伤掩码预测

第一阶段根据受损颜色体积预测精细化的体积损伤掩码。我们处理从 C^{dam}中提取的轴向 RGB 切片。对于每个切片索引 z，将 RGB 切片与根据受损体积生成过程计算出的粗略二值损伤指示图拼接，形成四通道输入：

$$x_z \in \mathbb{R}^{4\times32\times32}.$$

二维 U-Net 预测逐像素概率图 $\hat{} M_z \in [0,1]^{32\times32}$，并以 0.5 为阈值将其二值化，得到二值切片级掩膜。将所有切片的预测结果沿深度维度堆叠，并通过逻辑或运算进行合并。随后，使用一个包含$3\times3\times3$结构元素的小型三维形态学闭运算去除孤立间隙，从而生成最终的体积掩码

$$\widehat{M} \in \{0,1\}^{32\times32\times32}.$$

架构与训练。 二维掩码网络采用标准的 U-Net 架构，包含四个下采样阶段（64–512 个通道），并通过跳跃连接进行对称上采样。网络使用二元交叉熵损失和 Adam 优化器进行 50 个训练轮次的训练，学习率为 10^{-4}。数据增强包括随机翻转和平面内旋转，旋转角度最大为 $\pm15°$。

3.3 阶段 2：掩码条件扩散修复

第二阶段根据预测的损伤掩码重建缺失几何和颜色。掩码输入计算如下：

$$V^{\text{mask}} = V^{\text{dam}} \odot (1-\widehat{M}), \qquad C^{\text{mask}} = C^{\text{dam}} \odot (1-\widehat{M}),$$

其中 $\odot$ 表示逐元素乘法。

扩散模型的输入是一个 5 通道体素张量

$$X = [\,V^{\text{mask}} \parallel \widehat{M} \parallel C^{\text{mask}}\,] \in \mathbb{R}^{5\times32\times32\times32}.$$

我们采用去噪扩散概率模型，并在 $T = 1000$个时间步上使用线性噪声调度。在每次训练迭代中，均匀采样一个时间步 t，并使用注入网络瓶颈的正弦嵌入对其进行编码。

网络架构。 扩散主干网络是一个具有四个分辨率层级的三维 U-Net。每个编码器块由两个带有 ReLU 激活和残差连接的 $3\times3\times3$ 卷积组成，后接 $2\times2\times2$ 最大池化。通道宽度随层级增加，从 32 增至 128。解码器采用转置卷积和跳跃连接，镜像这一结构。两个独立的 $1\times1\times1$ 输出头分别预测占据对数几率体和 3 通道颜色残差。

最终占据情况通过应用 Sigmoid 函数并以 0.5 为阈值进行二值化得到。最终颜色值通过将预测的残差加到 C^{mask} 仅限掩膜区域内的颜色值上计算得出。

训练目标。 模型通过加权组合多种损失进行训练：(i) 扩散噪声预测损失（占据通道上预测噪声与目标噪声之间的均方误差），(ii) 对最终占据概率应用二元$\widehat{\ }$交叉熵，(iii) 在 M内部 RGB 值上的掩膜 L_1 损失，(iv) 使用 VGG-16 的 conv3_3之前各层特征计算的逐切片感知损失，以及 (v) 鼓励形成合理类陶瓷颜色分布的弱颜色先验。损失权重设为：占据 BCE 为 1.0，颜色 L_1为 20.0，感知损失和颜色先验项为 0.1。

优化。 训练使用 Adam 优化器进行 100 个训练轮次，初始学习率为 10^{-3}，批量大小为 4 个体积数据。使用因子为 0.5、耐心值为 5 的 ReduceLROnPlateau 调度器。以 0.5 的概率沿轴向进行随机镜像。

4 数据集与预处理

4.1 工件收集

我们整理了一个包含 23 件瓷器风格工件的数据集，其中包括真实世界扫描和 CAD 模型。具体而言，该数据集包含 9 个来自史密森尼 3D 扫描收藏（公有领域）的带纹理三维扫描，以及 14 个从 Free3D 获取、采用知识共享许可的 CAD 模型。这些物体经过筛选，涵盖了各种形状复杂度，包括平滑、对称的形态和更精细复杂的装饰几何形状。

所有网格均经过人工检查，以移除不连通组件；随后使用 MeshLab 中的二次边折叠简化算法，将面数简化至约 100,000 个。数据集按物体划分为 16 个训练工件、4 个验证工件和 3 个测试工件，并确保各划分中的形状复杂度均衡。



图 1：精选文物数据集中的代表性带纹理网格。

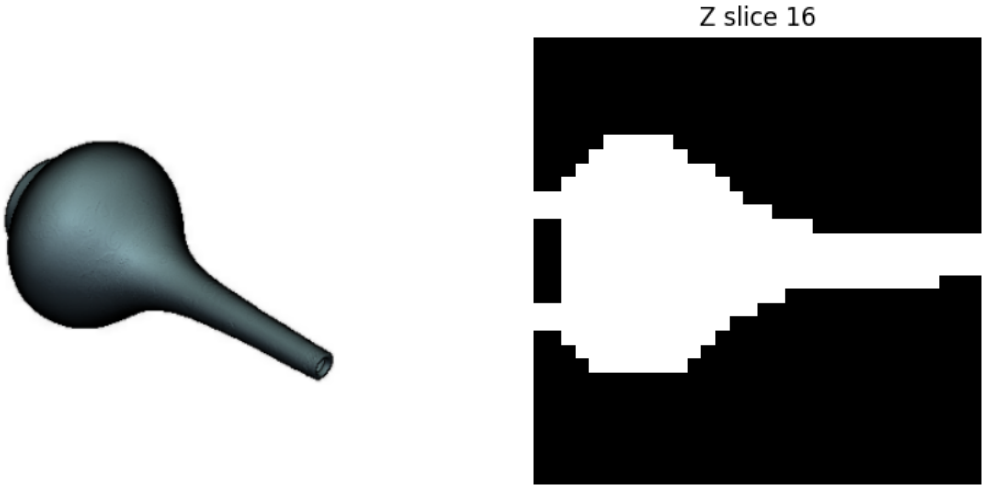


图 2：示例网格及其对应的 32^3 体素占据网格。

4.2 体素化与颜色体积

每个网格都会归一化，使其适配以原点为中心的单位立方体，并以固定分辨率 $32 \times 32 \times 32$ 进行体素化。体素化采用光线投射方法，将中心与网格表面相交的体素标记为占用。尽管原始网格以毫米为单位定义，但所有物体在体素化之前都会进行统一归一化；因此，报告的距离对应一致的相对尺度，评估时则将一个体素视为一毫米。

随后从外部进行泛洪填充，以生成一个封闭的二值占用网格

$$V^{\mathrm{gt}} \in \{0,1\}^{32\times32\times32}.$$

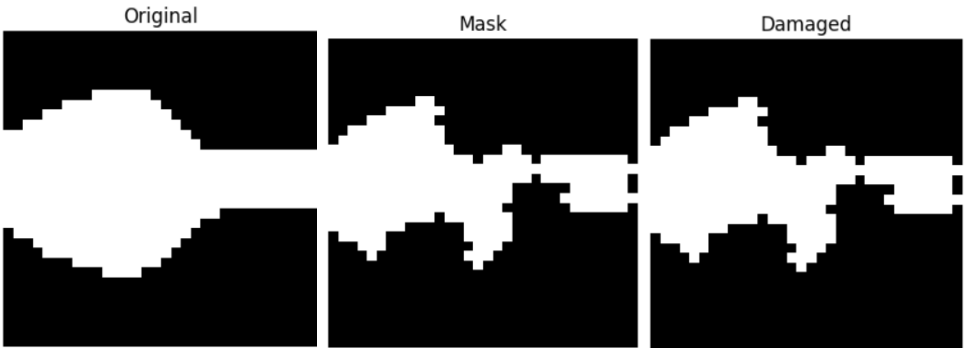
使用 Blender 从原始带纹理网格渲染每个体素的 RGB 颜色值，并将其与体素网格对齐，从而得到颜色体积

$$C^{\mathrm{gt}} \in [0,1]^{3\times32\times32\times32}.$$

颜色值仅为占据体素定义；未占用体素被赋予零颜色 r.

4.3 合成损伤生成

为了生成用于修复的训练对，我们直接对体素化的真实体积数据施加合成损伤。对于每个轴向切片，随机采样一至三个孔洞区域。这些区域为圆形或多边形，半径在 5 至 10 个体素之间均匀抽取，并从相应切片的占用网格中移除。



(a) 完整切片 (b) 损伤掩码 (c) 受损切片

图 3：展示完整几何结构、合成损伤掩码和受损输入的轴向体素切片示例。

逐切片孔洞移除后，使用半径为两个体素的球形结构元素对整个体积数据进行三维二值腐蚀，以模拟表面磨损和崩缺。此过程生成受损的占用网格

$$V^{\mathrm{dam}} \in \{0,1\}^{32\times 32\times 32},$$

以及一个二值损伤掩码

$$M = \mathbb{1}\left[V^{\mathrm{gt}} = 1 \ \wedge \ V^{\mathrm{dam}} = 0\right].$$

受损颜色体积通过对真实颜色体积应用掩码计算得出：

$$C^{\mathrm{dam}} = C^{\mathrm{gt}} \odot V^{\mathrm{dam}},$$

其中 $\odot$ 表示逐元素乘法。

4.4 归一化与数据增强

所有 RGB 值都归一化到范围 $[0,1]$。在第一阶段掩码预测训练中，使用 ImageNet 均值和标准差对 RGB 切片进行标准化，这与预训练编码器的初始化保持一致。在第二阶段扩散训练中，体素占据和 RGB 值在 $[0,1]$ 中直接使用，不再进行额外归一化。

第一阶段的数据增强包括随机水平翻转和垂直翻转，以及最大可达 $\pm 15^{\circ}$ 的平面内旋转。第二阶段以 0.5 的概率沿轴向进行随机镜像，以促进对称性感知泛化。

4.5 数据集大小与采样

每件文物可产生 32 个用于掩码预测的轴向切片，因此每个训练轮次包含 512 个训练切片。对于基于扩散的修复，每件文物在每个训练轮次都会生成一个独立受损的体积。因此，每个训练轮次包含 16 个体积样本；经过 100 个训练轮次后，扩散模型共使用 1,600 种不同的损伤实现进行训练。

5 实验、结果与讨论

我们在留出文物上评估所提出的两阶段流程，衡量几何补全准确度和颜色保真度。我们首先概述训练方案和超参数，然后定义评估指标，最后呈现定量和定性结果。

5.1 训练方案

阶段 1：损伤掩码预测。UNet2DColor 用于预测从体素化文物中提取的 32×32RGB 切片上的损伤掩码。我们使用 Adam 训练 50 个训练轮次，学习率为

训练轮次	扩散模型		对称性基线	
	倒角距离 ↓	F-score ↑	倒角距离 ↓	F-score ↑
10	0.0052	0.762	0.0123	0.512
50	0.0039	0.815	0.0110	0.533
80	0.0031	0.846	0.0106	0.545
100	0.0032	0.842	0.0105	0.548

Table 1：扩散模型与仅对称基线的验证几何指标。越低 r Cha越好，F-score 越高越好。

训练轮次	掩码均方误差 ↓	峰值信噪比（dB）↑
10	0.00345	24.62
50	0.00221	26.56
80	0.00198	27.03
100	0.00205	26.89

Tabl第 2：重叠占据体素上的验证颜色指标。较低的均方误差和较高的峰值信噪比表示更好 d的结果

1×10^{-4} 和 $(\beta_1,\ \beta_2) = (0。9,\ 0。999)$，批量大小为 8 个切片。损失函数为二元交叉熵。数据增强包括随机水平/垂直翻转，以及最大 $\pm 15°$ 的平面内旋转。我们选择验证损失最低的检查点。

阶段 2：基于扩散的体素修复。VoxelInpaintUNet 通过扩散过程进行训练，以在 32^3 分辨率下联合重建占据情况和颜色。我们使用 Adam 训练 100 个训练轮次，初始学习率为 1×10^{-3}，并采用 ReduceLROnPlateau 调度器（因子为 0.5，耐心值为 5）。批量大小为 4 个体积数据，输入形状为 $5 \times 32 \times 32 \times 32$。我们在 $T = 1000$ 个时间步上使用从 10^{-4} 到 2×10^{-2} 的线性 β 调度，并在每次迭代中随机采样一个时间步。数据增强包括以 0.5 的概率沿轴向进行镜像。模型选择基于验证损失和几何指标。

所有实验均在单个 NVIDIA A100 GPU 上运行。

5.2 评估指标

几何。我们使用倒角距离和 F-score，将预测的占据情况 $\widehat{V}$ 与真实值 V^{gt} 进行比较。倒角距离在占据体素集合之间计算，使用最近邻查询（通过 scipy.spatial.cKDTree 实现）。在假设归一化后一个体素对应一毫米的情况下，距离以毫米为单位报告。F-score 在 1 毫米阈值下计算，作为占据体素上精确率和召回率的调和平均值。

颜色。颜色保真度仅在 $\widehat{V}$ 和 V^{gt} 中均被占据的体素上进行评估。我们报告掩码均方误差和峰值信噪比，并额外计算各轴向切片的峰值信噪比，以分析空间变化。

5.3 定量结果

几何补全。 表 1 比较了扩散模型与仅对称基线的倒角距离和 F-score（1 mm）。在仅对称基线中，缺失体素通过将占据体素沿主要轴向对称平面镜像来填充。扩散模型在两项指标上均持续取得更好的表现，并在训练轮次约为 80 时达到最佳性能。

颜色重建。 表 2 报告了重叠占据体素上的掩码均方误差和峰值信噪比。性能随着训练稳步提升，并在训练轮次约为 80 时达到峰值。

5.4 定性结果

掩码预测。 图4展示了具有代表性的二维切片，其中将真实值和预测的损伤掩码叠加在 RGB 图像上。模型能够捕捉不规则的孔洞边界和局部缺失区域。



图4：第一阶段掩码预测：（左）RGB 切片，（中）真实掩码叠加图，（右）预测掩码叠加图。

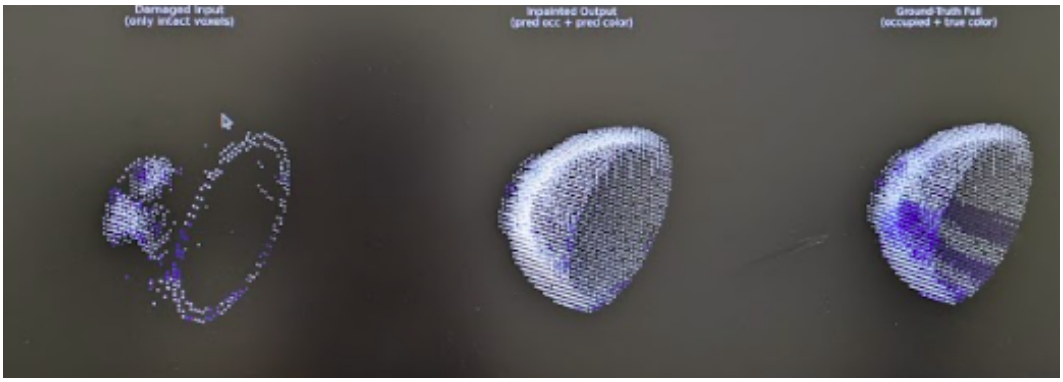


图5：具有鲜明图案的文物三维修复：受损输入与扩散重建及真实值的对比（按图中方式渲染）。

三**维修复**。 图5和图6展示了两个具有代表性的重建结果。与受损输入相比，扩散模型能够恢复缺失结构，并生成连贯的颜色过渡。

Per-切片峰值信噪比。图7绘制了性能最佳的检查点在各轴向切片上的峰值信噪比（训练轮次 h 80). 大多数切片的峰值信噪比超过 25 dB，但包含精细装饰图案的区域数值有所下降。

5.5 讨论与局限性

在留出文物上，掩码条件控制与基于扩散的体素修复相结合，相比仅对称基线取得了显著改进。在 32^3 体素化条件下，几何补全效果出色，能够可靠地恢复缺失结构。颜色修复总体上较为合理，但对纹理复杂度和有限的训练多样性仍较为敏感。具有高度精细或非对称表面图案的文物最具挑战性；在这些情况下，即使几何形状正确，重建颜色仍可能偏离真实值。

未来的工作可以通过引入更强的外观先验、更高分辨率的表示或多视图图像条件来提升颜色保真度。

6 结论与未来工作

本研究提出了一种轻量级框架，用于对体素化三维物体进行几何与颜色联合修复，旨在实现受损文化遗产文物的数字修复。通过将损伤定位与重建解耦，所提出的流程使基于扩散的模型能够专注于填补缺失区域，同时保留现有结构。在一组规模较小但

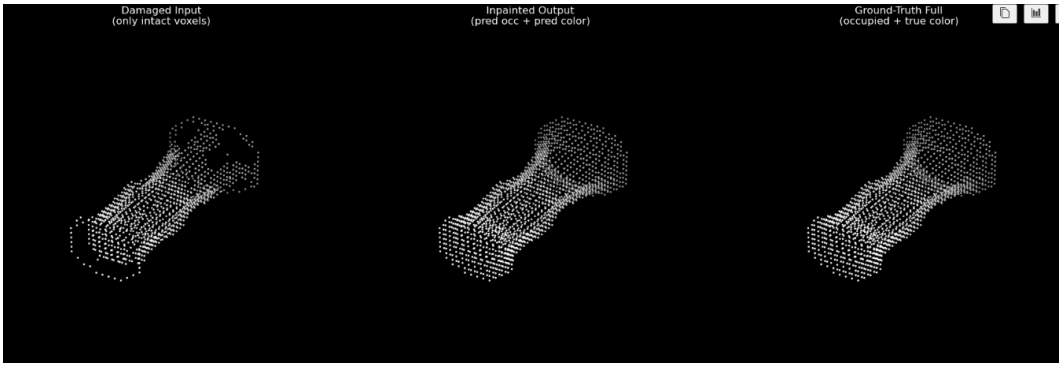


图 6：表面近乎均匀的文物的三维修复：受损输入、扩散重建与真实值（如图所示的渲染结果）。

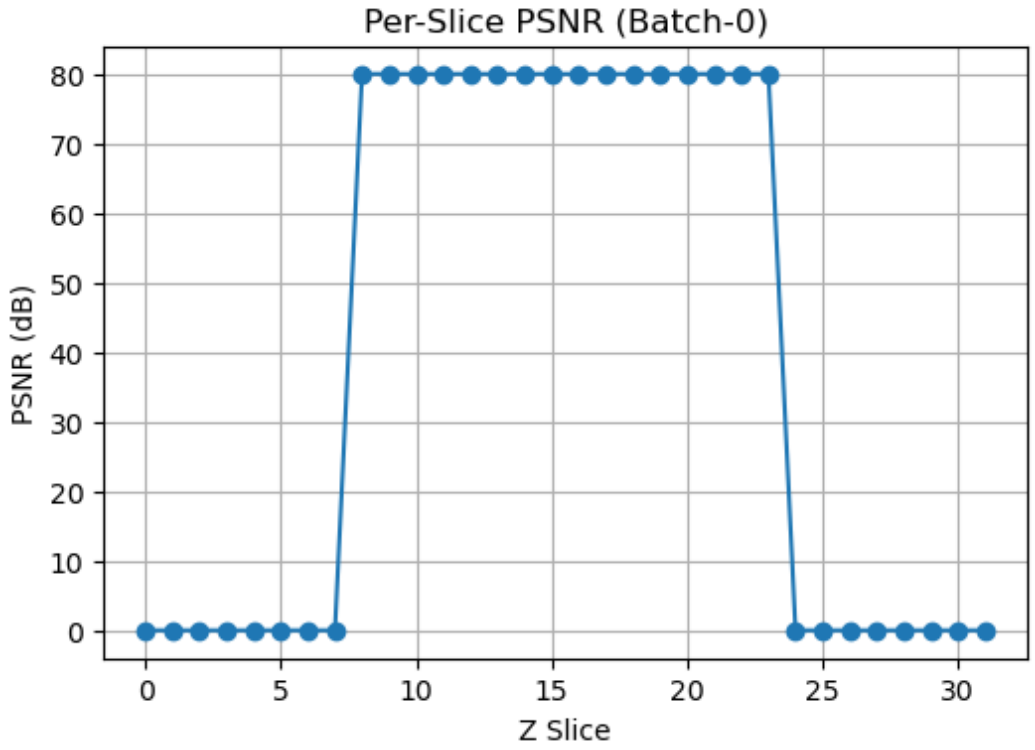


图 7：训练轮次 80 时验证集上的逐切片峰值信噪比。

种类多样的带纹理文物上进行的实验表明，在我们的实验所采用的 32^3 分辨率下，显式的掩码条件控制相比基于对称性的基线方法，能显著提升几何补全效果。

尽管在所评估的分辨率下，几何重建始终准确，但颜色修复仍更具挑战性，尤其是对于表面图案不对称或细节极其丰富的文物。这些局限凸显了从有限数据和低分辨率体积表示中学习外观先验的困难。

未来可以从多个方向拓展这项工作。采用更高分辨率的体素网格或混合体素—隐式表示，或许能提升几何和颜色保真度。融入更丰富的外观线索，例如多视角照片、表面法线或学习得到的纹理先验，或可进一步提升颜色重建的稳定性。最后，交互式或用户引导的修补机制，可能在实际修复工作流程中为自动化与专家控制提供实用的平衡。

参考文献

Christopher B. Choy、Danfei Xu、JunYoung Gwak、Kevin Chen 和 Silvio Savarese。3d-r2n2：一种用于单视角和多视角三维物体重建的统一方法。发表于欧洲计算机视觉会议（*ECCV*），2016年。Özgür Çiçek、Ahmed Abdulkadir、Stefan S. Lienkamp、Thomas Brox 和 Olaf Ronneberger。3D U-Net：从稀疏标注中学习密集体积分割。医学图像计算与计算机辅助介入（*MICCAI*），2016年。Angela Dai、Charles Ruizhongtai Qi 和 Matthias Nießner。Scancomplete：面向三维扫描的大规模场景补全与语义分割。发表于*IEEE*计算机视觉与模式识别会议（*CVPR*），2018年。Justin Johnson、Alexandre Alahi 和李飞飞。用于实时风格迁移和超分辨率的感知损失。发表于欧洲计算机视觉会议（*ECCV*），2016年。Michael Kazhdan 和 Hugues Hoppe。屏蔽泊松曲面重建。*ACM*图形学汇刊，32(3)，2013年。Michael Kazhdan、Matthew Bolitho 和 Hugues Hoppe。泊松曲面重建。发表于*Euro- graphics* 几何处理研讨会，2006年。Gimin Nam、Mariem Khlifi、Andrew Rodriguez、Alberto Tono、Linqi Zhou 和 Paul Guerrero。3D-LDM：基于潜在扩散模型的神经隐式三维形状生成。*arXiv* 预印本 *arXiv:2212.00842*，2022年。Michael Schröppel、Francis Engelmann 和 Matthias Nießner。Sc-diff：利用扩散模型进行三维形状补全。*arXiv* 预印本 *arXiv:2403.12470*，2024年。Karen Simonyan 和 Andrew Zisserman。用于大规模图像识别的超深卷积网络。*arXiv* 预印本 *arXiv:1409.1556*，2014年。Yaoqing Yang、Chen Feng、Yiru Shen 和 Dong Tian。Foldingnet：通过深度网格变形实现的点云自编码器。发表于*IEEE*计算机视觉与模式识别会议（*CVPR*），2018年。Wentao Yuan、Tejas Khot、David Held、Christoph Mertz 和 Martial Hebert。PCN：点云补全网络。*arXiv* 预印本 *arXiv:1808.00671*，2018年。