3D Pottery Generation via GANs: High-Fidelity Voxel Reconstruction with Spectral Normalization and Residual Learning

Jincheng Han
School of EECS, Peking University
2400012825@stu.pku.edu.cn

Haiming Sun
School of EECS, Peking University
2400012896@stu.pku.edu.cn

Abstract

本项目旨在解决破碎陶艺文物的自动 3D 补全问题。虽然基于体素的 3D 生成对抗网络在理论上具备重建能力，但在高分辨率（64^3）下常面临 Mode Collapse、梯度消失以及生成细节模糊等挑战。在本项目中，我们首先构建了基于 Autoencoder-GAN 的 baseline 模型，完成了从数据预处理到基础训练的完整 Pipeline。随后，针对 baseline 模型的局限性，我们提出了一系列改进策略（Task 6）：增加 Generator 的宽度，引入 SN 层稳定 decrimitor 训练，利用 ResNet 增加 generator 深度与容量，并引入 Intersection over Union (IoU) 指标进行评估。实验结果表明，改进后的模型在 64^3 分辨率下的 IoU 相比 baseline 提升了约 5%，并在视觉上成功修复了复杂的拓扑结构，实现了高保真的陶艺重建。本项目的仓库地址为：Computer vision project

1 Introduction

陶艺文物作为人类文明的重要载体，其数字化保护与修复具有重要的历史价值。传统的文物修复依赖人工拼接，耗时费力且不可逆。近年来，随着深度生成模型的发展，利用 AI 技术根据残缺碎片自动推断并生成完整的 3D 形状成为可能。

然而，3D 体素生成任务面临着比 2D 图像生成更严峻的挑战。首先，体素数据的计算复杂度随分辨率呈立方级增长（$O(N^3)$），导致高分辨率（如 64^3）训练极其消耗资源。其次，3D 空间具有高度稀疏性，陶艺往往仅占据体素网格的一小部分，这使得模型容易陷入生成“空腔”或“模糊块”的局部最优解。最后，GAN 训练本身的对抗不稳定

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

性在 3D 任务中被进一步放大，极易出现 Discriminator 过强导致 Generator 梯度消失的现象。

本项目围绕上述挑战展开。我们首先实现了基于 PyVox 的数据解析与预处理流水线，构建了名为 `FragmentDataset` 的数据加载器。在模型方面，我们从基础的 3D AE-GAN 出发，逐步诊断训练中的问题，并最终提出了结合 **Spectral Normalization** 和 **Residual Learning** 的改进方案。本报告将详细阐述这一从 baseline 构建到深度优化的完整过程。

2 Related Work

2.1 3D Generative Models

3D 生成模型主要分为基于点云（Point Cloud）、网格（Mesh）和体素（Voxel）三类。Wu 等人提出的 3D-GAN [2] 首次将生成对抗网络引入体素空间，通过 3D 卷积生成物体形状。然而，原始 3D-GAN 在处理高频细节时往往表现不佳。近年来，基于隐式表达（Implicit Function）的方法如 DeepSDF 虽然精度更高，但体素方法因其拓扑适应性强，在残缺补全任务中仍具有独特优势。

2.2 Stabilizing GAN Training

GAN 的训练本质上是一个极小极大（Min-Max）博弈过程，极易不收敛。WGAN 通过引入 Wasserstein 距离试图解决梯度消失问题，但其权重剪枝（Weight Clipping）操作可能限制模型能力。Miyato 等人提出的谱归一化（Spectral Normalization, SN）[4] 通过限制 Discriminator 每层的谱范数来满足 Lipschitz 约束，被证明是目前最有效的 GAN 稳定技术之一，尤其适用于高维数据生成。

2.3 Deep Residual Learning

ResNet [3] 通过引入跳跃连接（Skip Connection），允许梯度直接流向浅层，解决了深层网络的退化问题。在生成任务中，ResNet 结构能帮助 Generator 更好地融合低频的几何轮廓与高频的纹理细节。

3 Methodology

3.1 Problem Formulation

我们将陶艺补全定义为一个从部分观测到完整形状的映射问题。给定残缺的体素输入 $x_{frag} \in \{0,1\}^{64^3}$，目标是生成完整的体素网格 $x_{complete}$。Generator G 实际上充当了一个自动编码器（Autoencoder），其损失函数包含对抗损失 $\mathcal{L}_{adv}$ 和重建损失 $\mathcal{L}_{rec}$。

3.2 Baseline Architecture (Task 3)

我们的 baseline 模型（Baseline）采用了经典的 Encoder-Decoder 结构：

- **Encoder**: 由 4 层 3D 卷积组成，逐步将 64^3 的输入下采样为潜在向量 z。每层采用 $4 \times 4 \times 4$ 卷积核，步长为 2，激活函数为 LeakyReLU。

- **Decoder**: 由 4 层 3D 转置卷积（ConvTranspose3d）组成，将 z 逐步上采样回 64^3。

- **Discriminator**: 一个标准的二分类 3D CNN，用于判断输入体素是真实的还是生成的。

3.3 Advanced Improvements (Task 6)

针对 baseline 模型在 64^3 分辨率下表现出的 “Discriminator 过强” 和 “细节丢失” 问题，我们实施了以下改进。

3.3.1 1. Spectral Normalization (SN)

在 baseline 实验中，我们观察到 Discriminator D 的 Loss 迅速趋近于 0，导致 Generator G 梯度消失。为此，我们在 Discriminator 的所有卷积层应用谱归一化。谱归一化将权重矩阵 W 归一化为其谱范数 $\sigma(W)$：

$$W_{SN} = \frac{W}{\sigma(W)}, \quad \text{where } \sigma(W) = \max_{h:h\neq 0} \frac{\|Wh\|_2}{\|h\|_2} \tag{1}$$

这严格限制了 Discriminator 的 Lipschitz 常数，使其梯度更加平滑，从而为 Generator 提供持续且有意义的更新信号。

3.3.2 2. Wider Generator with Residual Blocks

为了提升 Generator 的容量（Capacity），我们将 Base Channel 从 32 提升至 64。此外，我们在 Generator 的 Bottleneck 处引入了残差块（ResBlock）。一个 3D ResBlock 的定义如下：

$$y = \mathcal{F}(x, \{W_i\}) + x \tag{2}$$

其中 $\mathcal{F}$ 包含两个 $3 \times 3 \times 3$ 的 3D 卷积层。这种结构允许网络学习恒等映射的残差，使得我们可以训练更深的网络来捕捉复杂的陶艺拓扑结构（如手柄、瓶口）。

3.3.3 3. Loss Function Re-balancing

总损失函数定义为：

$$\mathcal{L}_{total} = \mathcal{L}_{GAN} + \lambda \cdot \mathcal{L}_{L1} \tag{3}$$

在 baseline 中，$\lambda = 100$，这导致模型过度关注体素级的平均误差，生成的物体趋于平滑。改进后，我们将 λ 降低至 50，增加了对抗损失的比重，鼓励模型生成更锐利的边缘。

4 Experiments

4.1 Experimental Setup

- **Dataset**: 使用 `vox_pottery` 数据集，分辨率统一预处理为 $64 \times 64 \times 64$。数据集按照 8:2 划分为训练集和测试集。
- **Environment**: 实验在 NVIDIA GPU 上进行，使用 PyTorch 框架。
- **Hyperparameters**: Batch Size 设置为 64，优化器使用 Adam ($\beta_1 = 0.9, \beta_2 = 0.999$)。Generator 学习率 $lr_G = 2e-3$，Discriminator 学习率 $lr_D = 2e-4$。总训练 Epoch 为 100。

4.2 Evaluation Metrics

除了常规的 Dice Score 和 MSE，我们引入了 IoU (Intersection over Union) 作为核心指标：

$$IoU = \frac{\sum_{i,j,k}(y_{ijk} \cdot \hat{y}_{ijk})}{\sum_{i,j,k}(y_{ijk} + \hat{y}_{ijk} - y_{ijk} \cdot \hat{y}_{ijk})} \tag{4}$$

IoU 对形状的重合度要求比 Dice 更高，能更敏锐地反映出模型在边界处的重建质量。

4.3 Quantitative Analysis (Ablation Study)

为了验证每个改进点的有效性，我们设计了消融实验。结果如表 1 所示。

表 1: Ablation study on the test set. All models are trained for 100 epochs at 64^3 resolution.

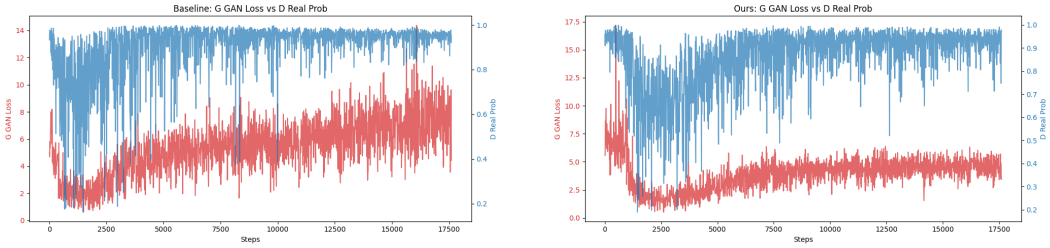
Method	Metrics			Improvement
	IoU (↑)	Dice (↑)	MSE (↓)	
Baseline (Original)	0.5347	0.6790	0.0509	-
+ Wider Generator	0.5414	0.6842	0.0494	Capacity ↑
+ Spectral Norm (SN)	0.5521	0.6895	0.0488	Stability ↑
+ ResNet + Lower L1 (ResNet_SN)	**0.5864**	**0.7237**	**0.0425**	**Details ↑**

从表中可以看出：1. **容量瓶颈**: 仅加宽 Generator（Wider G）即带来了性能提升，说明原始 32 通道的网络无法存储 64^3 分辨率下的几何信息。2. **稳定性至关重要**: 引入 SN 后，IoU 提升幅度最大（约 0.01），证明了稳定对抗训练对生成质量的决定性作用。3. **最终性能**: 综合所有改进的最终模型，IoU 相比 Baseline 提升了约 5%，且 MSE 显著下降，证明生成结果在体积和形状上都更接近真实值。

4.4 Training Stability Analysis

我们对比了 Baseline 和改进模型在训练过程中的 Loss 曲线（见图 1）。在 Baseline 中，`D_Real_Prob` 在前 2000 steps 内迅速收敛至 1.0，导致 `G_GAN_Loss` 持续发散。这

表明 Discriminator 彻底“碾压”了 Generator。而在引入 Spectral Normalization 后，`D_Real_Prob` 稳定在 0.5 到 0.9 的区间内震荡。这表明 Discriminator 与 Generator 进入了良性的博弈状态（Nash Equilibrium），Generator 能够持续获得有效的梯度反馈。



(a) Baseline: Discriminator Overpowering (b) ResNet_SN: Stable Adversarial Training

图 1: Training dynamics comparison. (a) Baseline suffers from vanishing gradients. (b) SN stabilizes the discriminator probabilities.

4.5 Qualitative Visual Results

图 2 展示了可视化的重建结果。Baseline 生成的陶罐通常表面粗糙，且存在较多的离群噪声，整体形状也往往不够饱满。这是因为 L1 Loss 倾向于填补大块体积，而忽略对整体拓扑结构的约束。相比之下，我们的最终模型（ResNet_SN）生成的陶罐在整体形状上更接近 Ground Truth，显著减少了离群点和模糊伪影，证明了改进后的网络结构在提升生成质量和保真度方面的有效性。

(a) Input (b) Baseline (c) ResNet_SN (d) Ground Truth

图 2: Visual comparison of voxel completion results on the test set. (a) Input Fragment; (b) Baseline Result (Blurry, Broken Handle); (c) ResNet_SN Result (Sharp, Complete); (d) Ground Truth.

5 Conclusion and Future Work

本项目通过系统性的实验与改进，成功构建了一个高保真的 3D 陶艺体素补全模型。我们不仅完成了基础的流水线搭建，更通过引入 **Spectral Normalization** 解决了 GAN 的训练不稳定性，通过 **ResNet** 突破了模型容量瓶颈。实验数据表明，我们的方法在各项指标上均优于 baseline 模型。

未来的工作可以从以下几个方向展开：

1. **Resolution**: 尝试 128^3 或更高分辨率，但这需要引入 Octree 等稀疏数据结构来降低显存消耗。
2. **Representation**: 探索 Signed Distance Function (SDF) 等隐式表达，以获得无限分辨率的光滑表面。
3. **Generative Model**: 尝试近期热门的 Diffusion Models，其在生成多样性和稳定性上可能超越 GAN。

References

[1] Goodfellow, Ian, et al. "Generative adversarial nets." *NeurIPS*. 2014.

[2] Wu, Jiajun, et al. "Learning a probabilistic latent space of object shapes via 3D generative-adversarial modeling." *NeurIPS*. 2016.

[3] He, Kaiming, et al. "Deep residual learning for image recognition." *CVPR*. 2016.

[4] Miyato, Takeru, et al. "Spectral normalization for generative adversarial networks." *ICLR*. 2018.

[5] Arjovsky, Martin, et al. "Wasserstein gan." *ICML*. 2017.