

1. はじめに

自動運転システムの閉ループ型評価を目的として、実世界シーンを再現し、任意視点の映像を生成する研究が行われている。中でも 3D Gaussian Splatting (3DGS) は、シーンを多数の 3 次元ガウス分布により表現し、写実性の高い映像を生成できる。3DGS の代表的な手法として Periodic Vibration Gaussian (PVG) [1] が提案されている。PVG は学習時に観測されていない新規視点の生成時、アーティファクトが生じやすいという課題がある。そこで本研究では、新規視点の生成時に、高品質な疑似生成画像を学習にフィードバックすることで、生成した新規視点の品質向上を目指す。

2. 関連研究

3DGS を時間方向に拡張した PVG とレンダリング画像を高品質化させる DIFIX について述べる。

2.1. Periodic Vibration Gaussian (PVG)

3DGS は 3 次元空間に、色や不透明度の 3 次元ガウス分布を分布を配置し、特定の視点からの画像を生成する。PVG[1] は、各 3 次元ガウス分布に時間表現を導入することで、動的シーンの映像生成を可能にしている。これにより、動的シーンにおいて時間的に滑らかな再構成が実現される。一方、新規視点においては視覚的品質が低下しやすく、アーティファクトが発生しやすい課題がある。

2.2. DIFIX

DIFIX[2] は、3DGS などの 3D 再構成手法によって生成された画像に含まれるアーティファクトを低減することを目的とした手法である。従来の多段階拡散モデルではなく、単一ステップの拡散モデルを用いることで、高速かつ実用的なアーティファクト除去を実現している。

3. 提案手法

本研究では、PVG による新規視点の生成画像の品質を向上させるために、DIFIX を用いて修復した生成画像を疑似教師画像として PVG の学習にフィードバックする手法を提案する。図 1 に提案手法の流れを示す。まず、実画像のみで 30,000 イテレーション学習した PVG を用いて新規視点の画像を生成する。次に、物体の一部が破綻しているかを目視で確認し、破綻していない生成画像のみを疑似教師画像とする。そして、元の視点の画像とともに PVG を学習する。これにより構造的破綻を学習することを防ぐ。

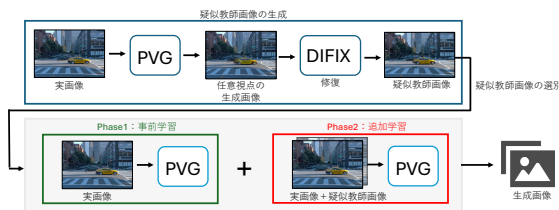


図 1: 提案手法の学習の流れ

4. 評価実験

追加する疑似教師画像の生成条件を変えて、PVG を学習した時の画像を比較し、提案手法の有効性を検証する。

4.1. 実験概要

図 2 に評価実験の学習の流れを示す。DIFIX による疑似教師画像の生成条件は以下の 3 つである。

- 1) 実画像のみ
- 2) 右へ 20cm 移動
- 3) 右へ 5~20cm 移動

条件 1 では、実画像のみを用いて 30,000 イテレーション学習した。条件 2 では、実画像で 15,000 イテレーション学習後、DIFIX で生成した新規視点画像を選別・追加し、15,000 イテレーションで学習した。条件 3 では、実画像で 15,000 イテレーション学習後、DIFIX で生成した新規視

点画像を選別・追加し、15,000 イテレーション学習した。各条件で学習した PVG を用いて右に 20cm ずらした新規視点画像の視覚的品質を比較する。評価には 4 シーンを用いる。

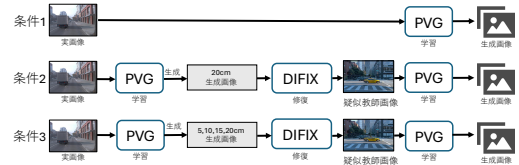


図 2: 各条件におけるレンダリング画像の生成

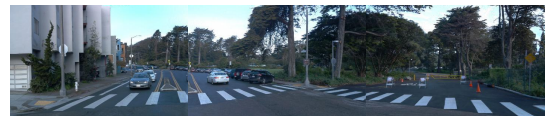
4.2. 実験結果

表 1 に 4 シーンの定量的評価を示す。表 1 より、条件 3 は条件 1 と比較して PSNR が約 0.7pt 向上し、SSIM も約 0.06pt 向上していることが分かる。また、条件 3 は条件 2 と比較して PSNR が約 2.8pt 向上し、SSIM が約 0.09pt 向上していることが分かる。

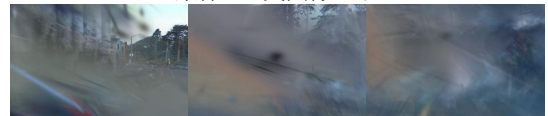
表 1: 4 シーンにおける平均値および分散

	条件 1 (実画像のみ)	条件 2 (20cm のみ)	条件 3 (5~20cm)
NIQE↓ (Mean)	3.284	3.525	3.585
NIQE↓ (Var)	0.108	0.120	0.116
PSNR↑ (Mean)	18.594	16.454	19.265
PSNR↑ (Var)	3.207	2.809	11.129
SSIM↑ (Mean)	0.497	0.471	0.556
SSIM↑ (Var)	0.00377	0.00539	0.00831
LPIPS↓ (Mean)	0.308	0.459	0.365
LPIPS↓ (Var)	0.00034	0.00013	0.00814

図 3 に 1 シーン目の定性的評価を示す。図 3 より、条件 2 と比較して条件 3 は、アーティファクトが含まれていないことが分かる。このことから、複数の軌跡の疑似教師画像を用いることで、新規視点画像における画素値の再現性および構造的な一貫性が改善されることが考えられる。しかし、条件 3 は条件 1 と比較して、白い車にアーティファクトを含んでいる。これより、提案手法による改善は不十分であることが分かった。



条件 1: 実画像のみ



条件 2: 疑似教師あり (20cm のみ)



条件 3: 疑似教師あり (5~20cm)

図 3: 定性的評価

5. おわりに

本研究では、DIFIX で修復した新規軌跡画像を疑似教師画像として PVG の学習にフィードバックする手法を提案した。評価実験より、条件 2 と比較して条件 3 の方が定性的・定量的評価で良いスコアが確認できた。今後は、より高品質な新規視点の疑似教師画像を生成することで、新規視点のレンダリング精度の向上に取り組む。

参考文献

- [1] Y. Huang *et al.*, “Periodic Vibration Gaussian for Dynamic Scene Reconstruction”, IJCV, 2026.
- [2] J. Z. Wu *et al.*, “DIFIX3D+: Improving 3D Reconstructions with Single-Step Diffusion Models”, CVPR, 2025.