ICCV23
PARIS
jeffreyxiang.github.io/GRAM-HD

GRAM-HD: 3D-Consistent Image Generation at High Resolution with Generative Radiance Manifolds

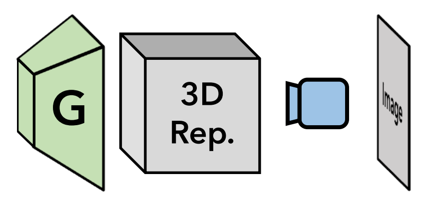
Jianfeng Xiang[1,2] Jiaolong Yang[2] Yu Deng[1,2] Xin Tong[2]
[1]Tsinghua University [2]Microsoft Research Asia

Project Page

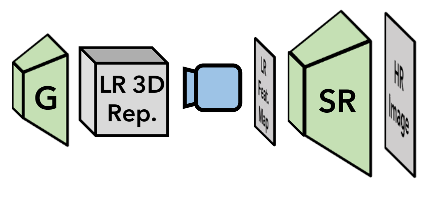
Tsinghua University
Microsoft

Overview

- **Task: 3D-aware Image Generation with GANs**
  Synthesizing the images of an object at different viewpoints, trained only on unstructured 2D images.
- **Current Problems**
  Prior works mainly following two different axes.

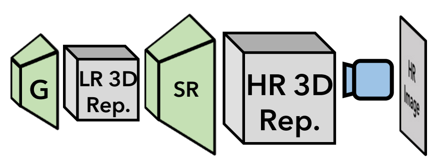


Pure 3D: The final images are directly rendered from output 3D.
✓ Strong consistency
× Limited resolution



3D+2D SR: Image-space SR is applied to get the final HR images.
✓ High resolution
× Poor consistency

- **Our Solution**
  We propose **GRAM-HD,** which gets the best of both worlds by employing 2D CNN for efficient **radiance manifold SR**.

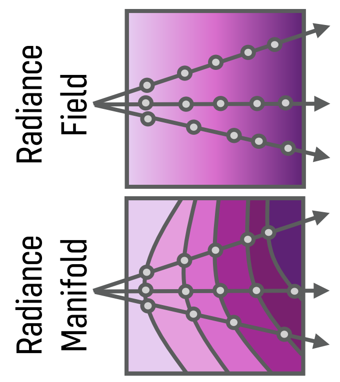


GRAM-HD: 3D SR with radiance manifold for HR images.
✓ **Strong 3D-consistentcy**
✓ **High resolution** results

- **Contributions**
  - o **Strongly Consistent high-resolution** 3D-aware image generation method
  - o **Efficient 3D super-resolution**

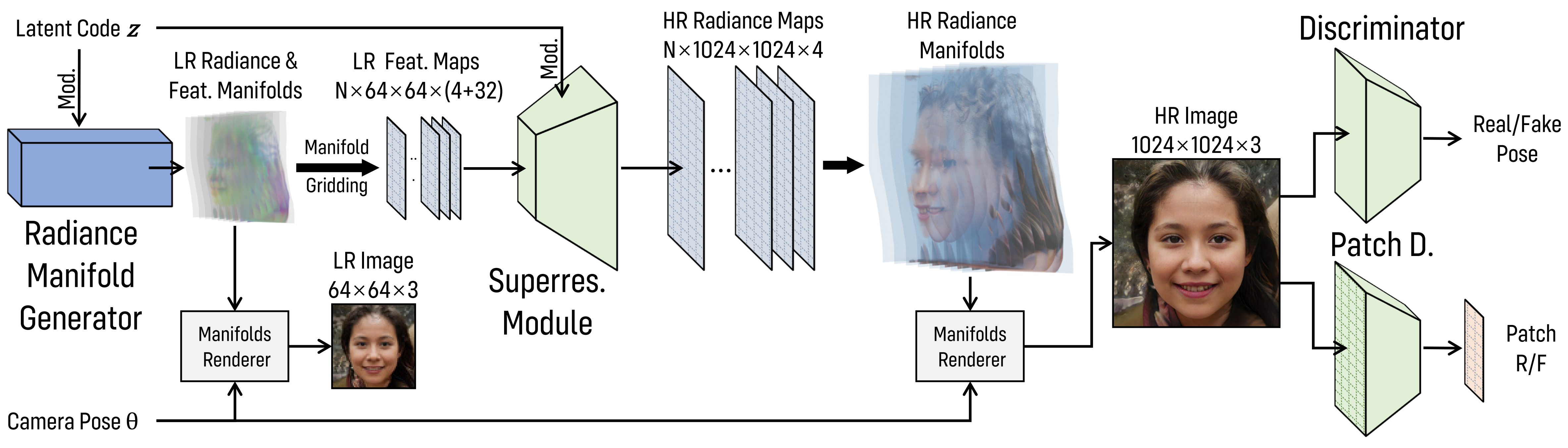
Methodology

- **Radiance Manifolds**
  The RF is defined in **continuous 3D space**. Rendering is **hard to afford.** Radiance manifolds regulate the sampling on **surface manifolds,** getting significantly **better efficiency.**



- **Manifold Super-resolution**
  For HR 3D manifolds generation, **efficient 2D CNNs** can be applied for **manifold super-resolution**.
- **Overall Pipeline**
  Two components: **manifold generator** & **manifold SR module**.
  - o Get **LR radiance and feature manifolds** representing LR 3D scenes from **manifold generator** .
  - o Flatten and discretize the manifolds to **LR 2D feature maps** through **manifold gridding**.
  - o Upsample to **HR radiance maps** with **manifold SR**.
  - o **HR images** can be rendered by integrating the HR radiance maps along ray-manifold intersections.
- **Network Training**
  A two-stage training strategy is employed.
  - o First, train LR model with **adversarial** & **pose loss**.
  - o Second, we train SR module with additional **patch adversarial** & **cross-resolution consistency loss**.

Framework



Results

Experiments

- **Quality Comparison**
  Ours has comparable quality to 3D+2D SR methods while superior to pure 3D methods.

Method	FFHQ256 FID	FFHQ256 KID	CATS256 FID	CATS256 KID
GRAM [13]	15.0	6.55	12.9	7.37
Ours	**11.8**	**4.72**	**7.05**	**2.53**

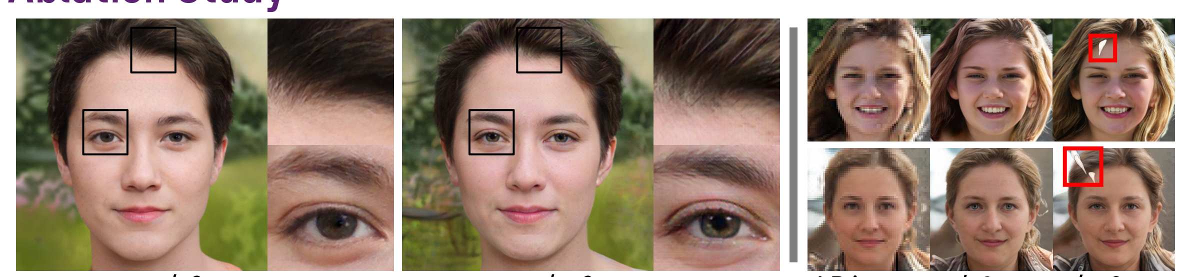
Method	FFHQ1024 FID	FFHQ1024 KID	FFHQ512 FID	FFHQ512 KID	CATS512 FID	CATS512 KID	FFHQ256 FID	FFHQ256 KID
StyleNeRF [20]	9.45	**2.65**	–	–	–	–	**9.24**	**3.19**
StyleSDF [48]	**9.44**	2.83	–	–	7.91	3.90	–	–
EG3D [7]	–	–	**8.72**	**3.61**	**6.28**	**1.67**	–	–
EpiGRAF [59]	–	–	14.7	6.93	9.40	4.11	16.1	7.73
Ours	12.0	5.23	12.2	5.41	7.67	3.41	11.8	4.72

- **Consistency Comparison**
  Consistency is compared with NeuS recon quality and EPI-akin images. Ours achieves superior multi-view consistency comparing to the competing methods.

Method	FFHQ1024 PSNR	FFHQ1024 SSIM	FFHQ512 PSNR	FFHQ512 SSIM	CATS512 PSNR	CATS512 SSIM	FFHQ256 PSNR	FFHQ256 SSIM
StyleNeRF [20]	30.0	0.80	–	–	–	–	31.9	0.92
StyleSDF [48]	31.1	0.84	–	–	26.6	0.75	–	–
EG3D [7]	–	–	33.7	0.88	28.4	0.78	–	–
Ours	**33.8**	**0.87**	**34.0**	**0.90**	**28.8**	**0.81**	**36.5**	**0.96**

StyleNeRF StyleSDF EG3D Ours

- **Ablation Study**



w/ $\mathcal{L}_{patch}$ w/o $\mathcal{L}_{patch}$ LR image w/ $\mathcal{L}_{cons}$ w/o $\mathcal{L}_{cons}$

Applications

- **High Resolution Image Embedding and Editing**
  By embedding an image into the latent space of trained model, pose editing can be conducted by rendering images at novel views. our method can generate high-fidelity novel view results.

Target Novel View Synthesis

- **Real-time Free-view Synthesis**
  By caching the manifolds and HR radiance maps, our method can support free-view video synthesis of **1024^2 images at real-time** frame rate.

Future Work

Our method still have several limitations.
× Handle objects with complex 3D geometry.
× Generation quality still lags 2D GANs.
Better representations or training strategies could be further explored to close the gap.