Generative Neural Scene Representations for 3D-Aware Image Synthesis

Michael Niemeyer

Autonomous Vision Group
University of Tübingen / MPI for Intelligent Systems Tübingen

Dec 7, 2021

University of Tübingen
MPI for Intelligent Systems
Autonomous Vision Group

Covered Papers

GRAF: Generative Radiance Fields for 3D-Aware Image Synthesis
Katja Schwarz and Yiyi Liao and Michael Niemeyer and Andreas Geiger
NeurIPS 2020

GIRAFFE: Representing Scenes as Compositional Generative Neural Feature Fields
Michael Niemeyer, Andreas Geiger
CVPR 2021

CAMPARI: Camera-Aware Decomposed Generative Neural Radiance Fields
Michael Niemeyer, Andreas Geiger
3DV 2021

Collaborators

Katja Schwarz

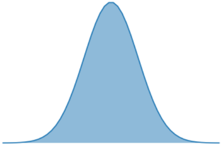
Yiyi Liao

Andreas Geiger

Generative Models are great!

Generative Models

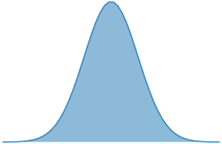
Sample a latent code from the prior distribution.



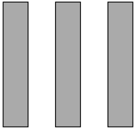
Latent Code

Generative Models

Pass latent code to trained generator G_θ.



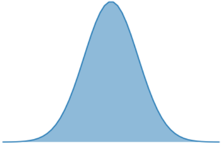
Latent Code



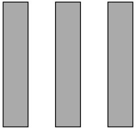
Generator G_θ

Generative Models

The generator outputs a synthesized image.



Latent Code



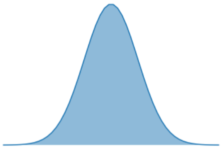
Generator G_θ

Generated Image*

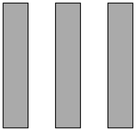
* The generated images are samples from StyleGAN2.

Generative Models

Sample more latent codes to get different generated images.



Latent Code



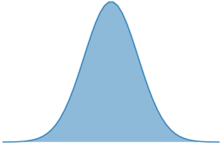
Generator G_θ

Generated Image*

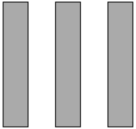
* The generated images are samples from StyleGAN2.

Generative Models

Sample more latent codes to get different generated images.



Latent Code



Generator G_θ

Generated Image*

* The generated images are samples from StyleGAN2.

Is the ability to sample photorealistic images
all we want?

Generative Models

For many applications, we require **control over the generation process**:

Generative Models

For many applications, we require **control over the generation process**:

Note: This and the following videos are only shown when opened with a supported PDF reader (e.g. Okular).

Animation Movies

Video Source: Disney's Toy Story 4 Trailer

Generative Models

For many applications, we require **control over the generation process**:

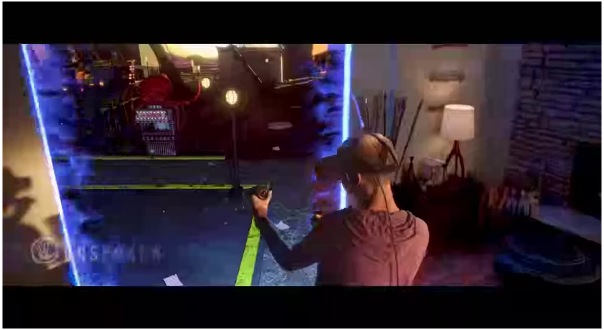
Video Games

Video Source: Gran Turismo 7 Trailer

Generative Models

For many applications, we require **control over the generation process**:

Virtual Reality



Video Source:Oculus Rift Trailer

Generative Models

Goal: A generative model for **3D-aware image synthesis** which allows us to:

Generative Models

Goal: A generative model for **3D-aware image synthesis** which allows us to:

- Generate photorealistic images

Generative Models

Goal: A generative model for **3D-aware image synthesis** which allows us to:

- Generate photorealistic images
- Control individual objects wrt. their pose, size, and position in 3D

Generative Models

Goal: A generative model for **3D-aware image synthesis** which allows us to:

- Generate photorealistic images
- Control individual objects wrt. their pose, size, and position in 3D
- Control camera viewpoint in 3D

Generative Models

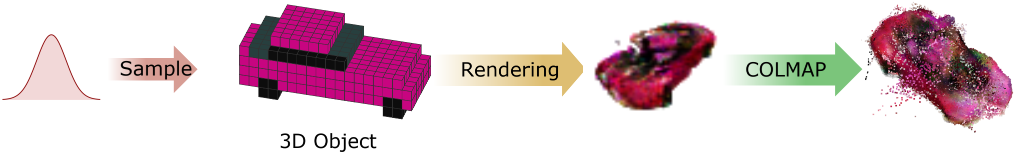
Goal: A generative model for **3D-aware image synthesis** which allows us to:

- Generate photorealistic images
- Control individual objects wrt. their pose, size, and position in 3D
- Control camera viewpoint in 3D
- Train from collections of unposed images

What representation should we use for
3D-aware image synthesis?

3D Representations

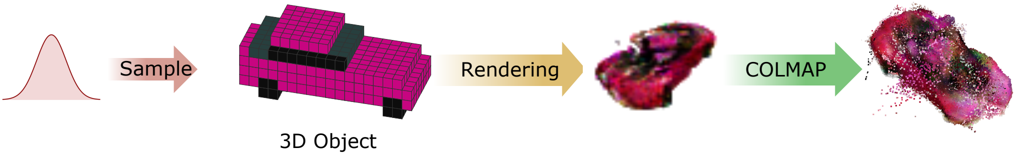
Voxel-based 3D Shape with Volumetric Rendering



PlatonicGAN [Henzler et al., ICCV 2019]

3D Representations

Voxel-based 3D Shape with Volumetric Rendering

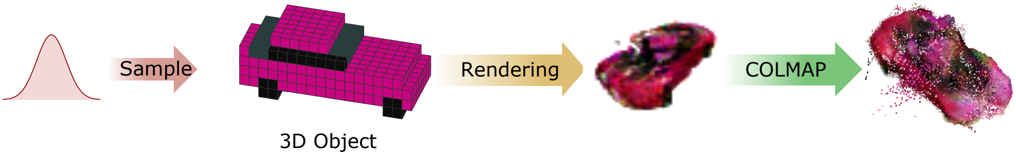


PlatonicGAN [Henzler et al., ICCV 2019]

+ Multi-view consistent

3D Representations

Voxel-based 3D Shape with Volumetric Rendering

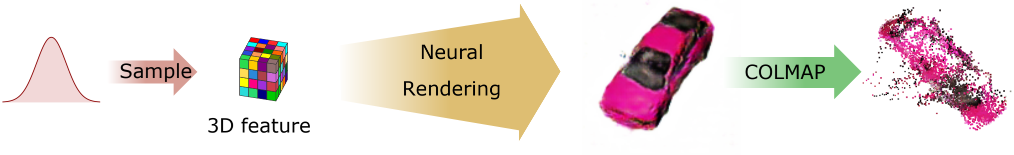


PlatonicGAN [Henzler et al., ICCV 2019]

+ Multi-view consistent
– Low image fidelity, high memory consumption

3D Representations

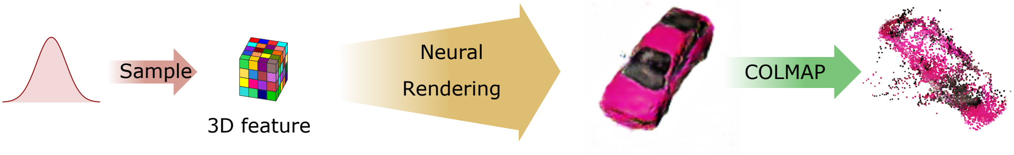
Voxel-based 3D Latent Feature with Learnable Projection



HoloGAN [Nguyen-Phuoc et al., ICCV 2019]

3D Representations

Voxel-based 3D Latent Feature with Learnable Projection

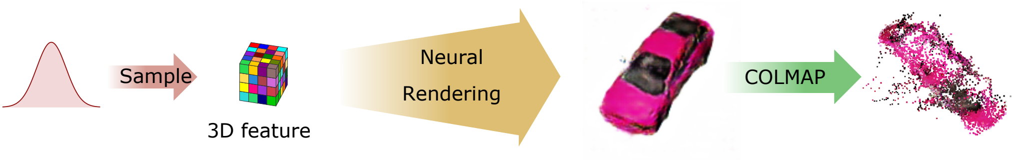


HoloGAN [Nguyen-Phuoc et al., ICCV 2019]

+ High image fidelity

3D Representations

Voxel-based 3D Latent Feature with Learnable Projection

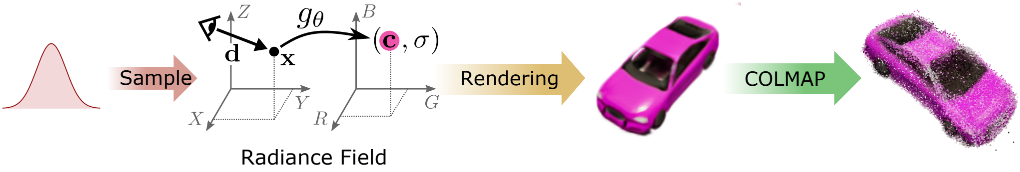


HoloGAN [Nguyen-Phuoc et al., ICCV 2019]

+ High image fidelity
– Object identity may vary with viewpoint due to learnable projection

3D Representations

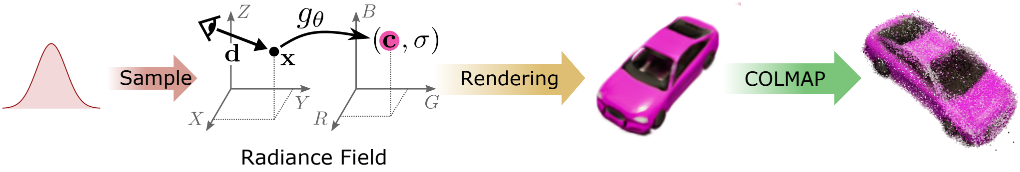
Generative Radiance Fields



Schwarz, Liao, Niemeyer, Geiger: GRAF: Generative Radiance Fields for 3D-Aware Image Synthesis. NeurIPS, 2020

3D Representations

Generative Radiance Fields

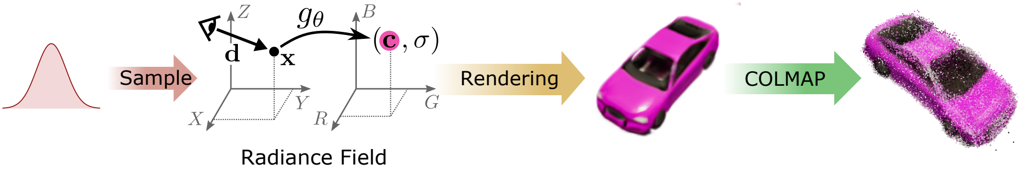


+ Continuous representation, multi-view consistent

Schwarz, Liao, Niemeyer, Geiger: GRAF: Generative Radiance Fields for 3D-Aware Image Synthesis. NeurIPS, 2020

3D Representations

Generative Radiance Fields



- ✚ Continuous representation, multi-view consistent
- ✚ High image fidelity, low memory consumption

Generative Radiance Fields

Generative Radiance Fields

Sample camera matrix $\mathbf{K}$, camera pose $\boldsymbol{\xi} \sim p_\xi$, and patch sampling pattern $\boldsymbol{\nu} \sim p_\nu$.

$\mathbf{K}$

$\boldsymbol{\xi} \sim p_\xi$

$\boldsymbol{\nu} \sim p_\nu$

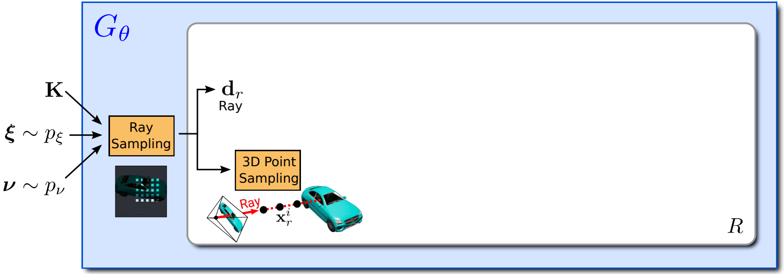
Generative Radiance Fields

Pass $\mathbf{K}$, $\boldsymbol{\xi}$, and $\boldsymbol{\nu}$ to generator G_θ and sample pixels / rays on image plane.

Schwarz, Liao, Niemeyer, Geiger: GRAF: Generative Radiance Fields for 3D-Aware Image Synthesis. NeurIPS, 2020

Generative Radiance Fields

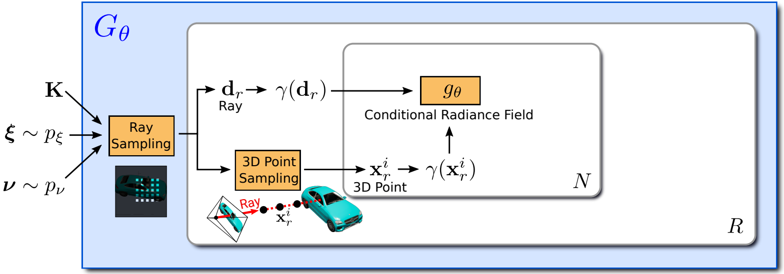
For each ray, get viewing direction $\mathbf{d}_r$ and sample 3D points $\mathbf{x}_r^i$ along ray.



Patch Generator

Generative Radiance Fields

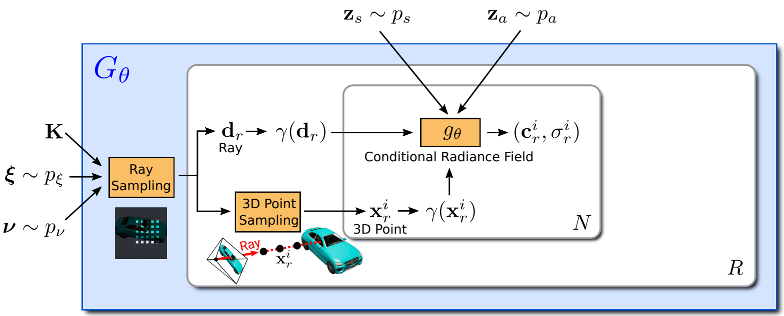
Pass $\mathbf{d}_r$ and $\mathbf{x}_r^i$ to positional encoding γ and then to the conditional radiance field g_θ.



Patch Generator

Generative Radiance Fields

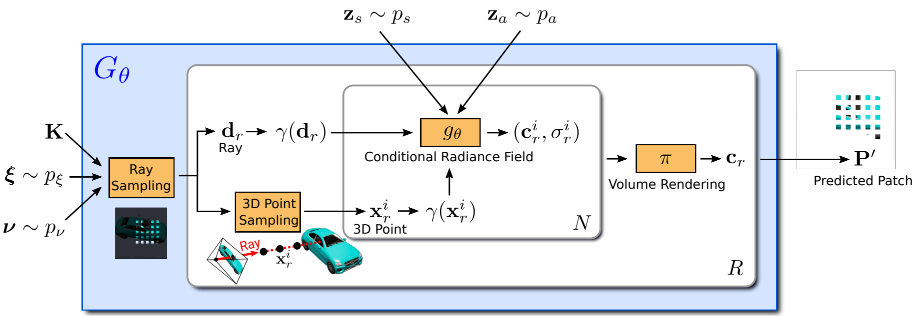
Sample latent shape and appearance codes $\mathbf{z}_s, \mathbf{z}_a$ and pass them to g_θ.



Patch Generator

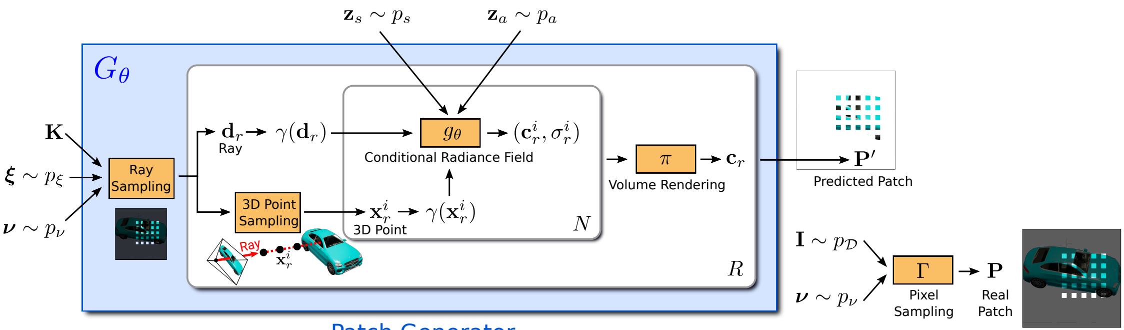
Generative Radiance Fields

Perform volume-rendering for each ray and get predicted patch $\mathbf{P}'$.



Generative Radiance Fields

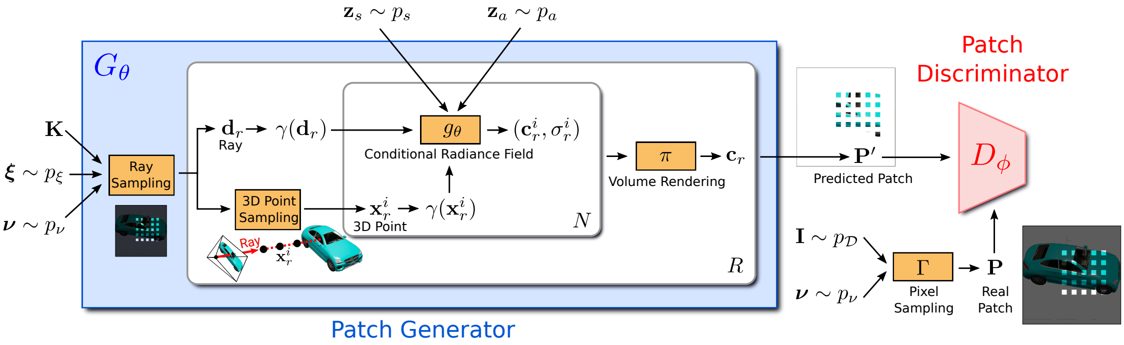
Sample patch $\mathbf{P}$ from real image $\mathbf{I}$ drawn from the data distribution $p_{\mathcal{D}}$.



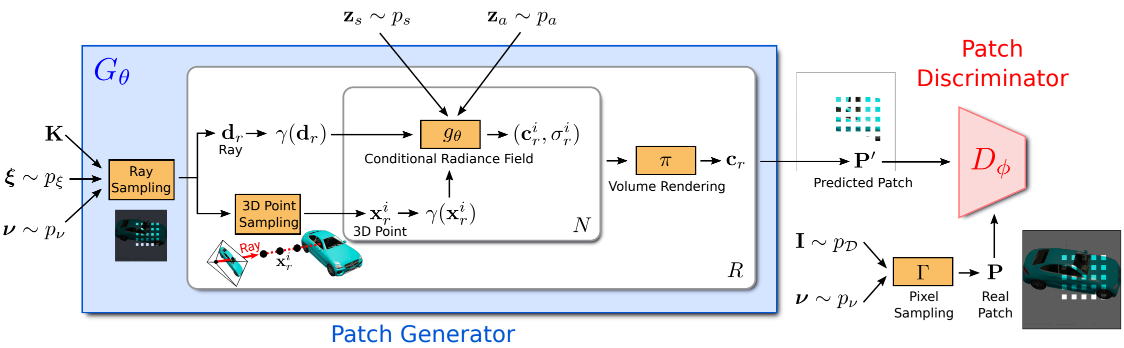
Patch Generator

Generative Radiance Fields

Pass fake and real patch $\mathbf{P}', \mathbf{P}$ to discriminator D_ϕ and train with adversarial loss.



Generative Radiance Fields



- Generator/discriminator for **image patches** of size 32 × 32 pixels
- Patches sampled at **random scale** using dilation

Volume Rendering

Rendering model for ray r(t) = o + td:

$$C \approx \sum_{i=1}^{N} T_i \alpha_i c_i$$

weights

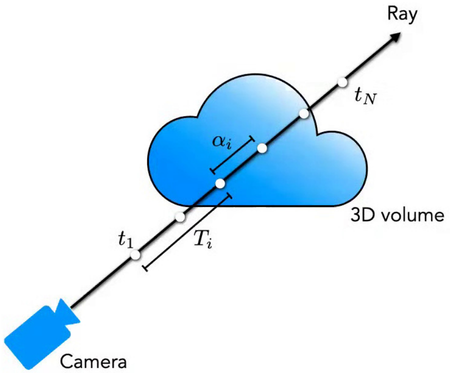
colors

How much light is blocked earlier along ray:

$$T_i = \prod_{j=1}^{i-1} (1 - \alpha_j)$$

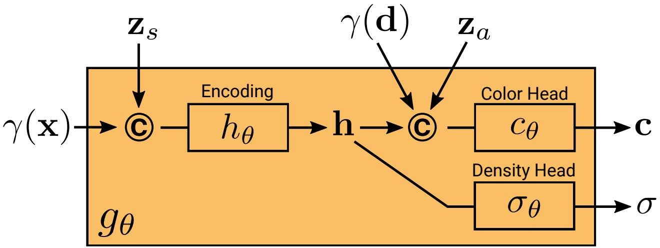
How much light is contributed by ray segment *i*:

$$\alpha_i = 1 - e^{-\sigma_i \delta t_i}$$



How do we parametrize
Conditional Radiance Fields?

Conditional Radiance Fields



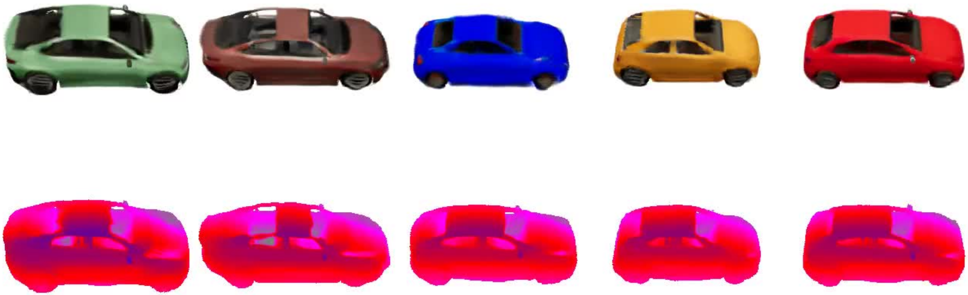
- Conditional radiance fields as fully-connected MLPs with ReLU activation
- Shape code $\mathbf{z}_s$ concatenated with encoded 3D location $\gamma(\mathbf{x})$
- Appearance code $\mathbf{z}_a$ concatenated with encoded viewing direction $\gamma(\mathbf{d})$

Schwarz, Liao, Niemeyer, Geiger: GRAF: Generative Radiance Fields for 3D-Aware Image Synthesis. NeurIPS, 2020

How well does it work?

Generative Radiance Fields

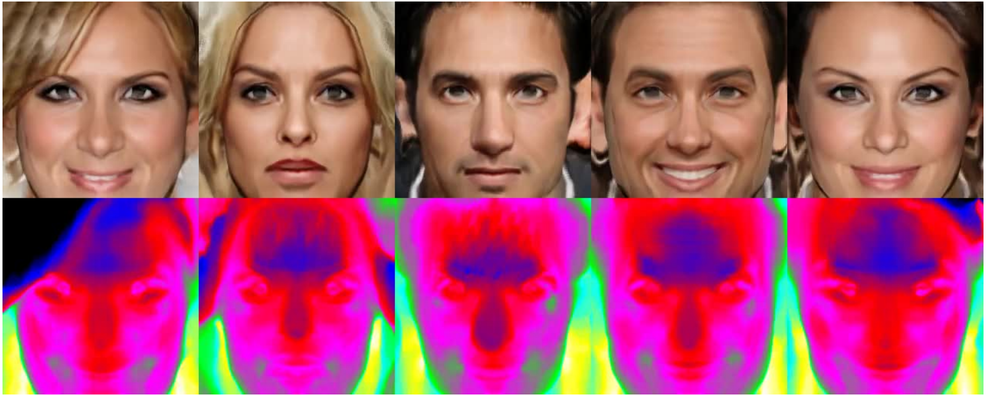
Results on synthetic Carla dataset at 256^2 pixels:



Schwarz, Liao, Niemeyer, Geiger: GRAF: Generative Radiance Fields for 3D-Aware Image Synthesis. NeurIPS, 2020

Generative Radiance Fields

Results on real CelebA-HQ dataset at 256^2 pixels:



Schwarz, Liao, Niemeyer, Geiger: GRAF: Generative Radiance Fields for 3D-Aware Image Synthesis. NeurIPS, 2020

How can we scale to
more complex, multi-object scenes?

GIRAFFE: Compositional Generative Neural Feature Fields

GRAF:

- Incorporate a **3D representation** into the generative model

GIRAFFE: Compositional Generative Neural Feature Fields

GRAF:

- Incorporate a **3D representation** into the generative model

GIRAFFE:

- Incorporate a **compositional 3D scene representation** into the generative model

Niemeyer, Geiger: GIRAFFE: Representing Scenes as Compositional Generative Neural Feature Fields. CVPR, 2021

GIRAFFE: Compositional Generative Neural Feature Fields

GRAF:

- Incorporate a **3D representation** into the generative model

GIRAFFE:

- Incorporate a **compositional 3D scene representation** into the generative model
- Incorporate a **neural renderer** to yield fast and high-quality inference

Niemeyer, Geiger: GIRAFFE: Representing Scenes as Compositional Generative Neural Feature Fields. CVPR, 2021

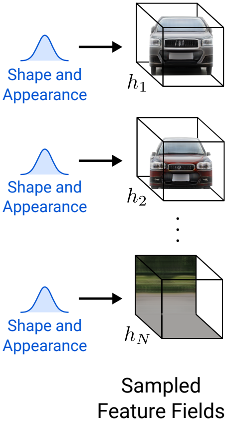
GIRAFFE

GIRAFFE

Sample N shape and appearance codes.

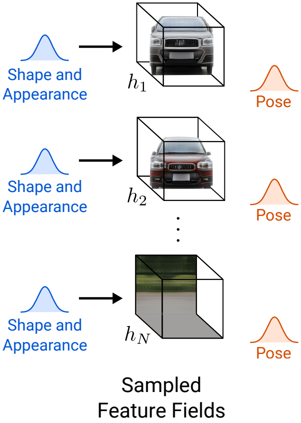
GIRAFFE

Get N feature fields. Note: We show features in RGB color for clarity.



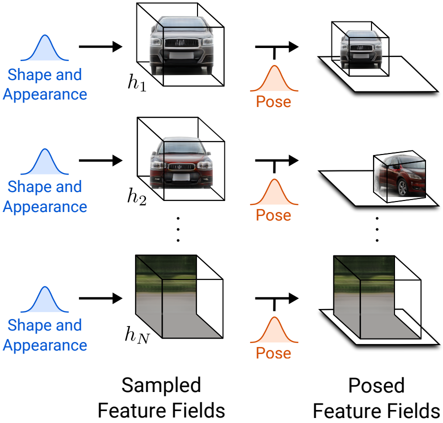
GIRAFFE

Sample size and pose for each feature field.



GIRAFFE

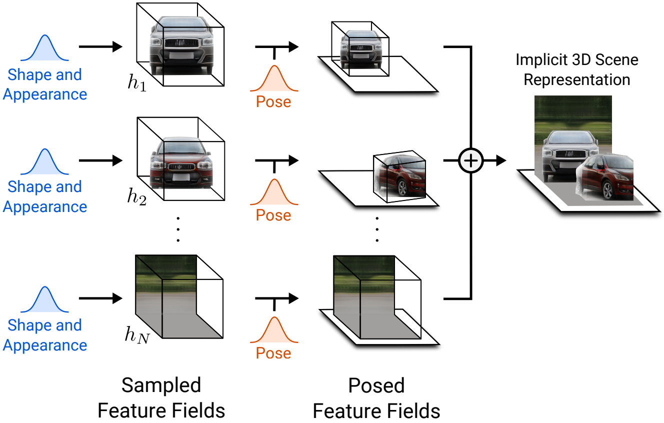
Get posed feature fields.



Niemeyer, Geiger: GIRAFFE: Representing Scenes as Compositional Generative Neural Feature Fields. CVPR, 2021

GIRAFFE

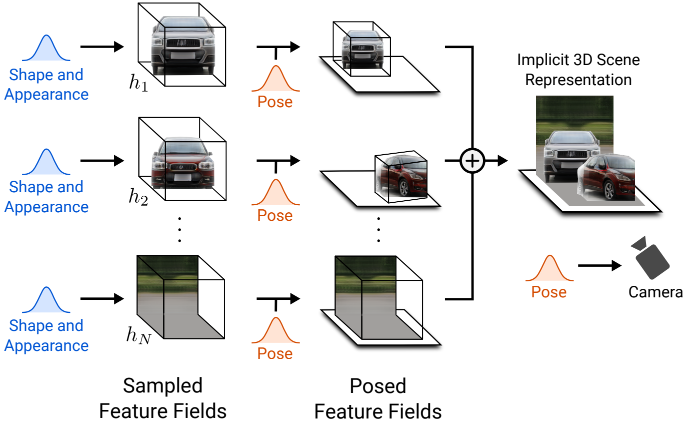
Composite all feature feature fields to one 3D scene representation.



Niemeyer, Geiger: GIRAFFE: Representing Scenes as Compositional Generative Neural Feature Fields. CVPR, 2021

GIRAFFE

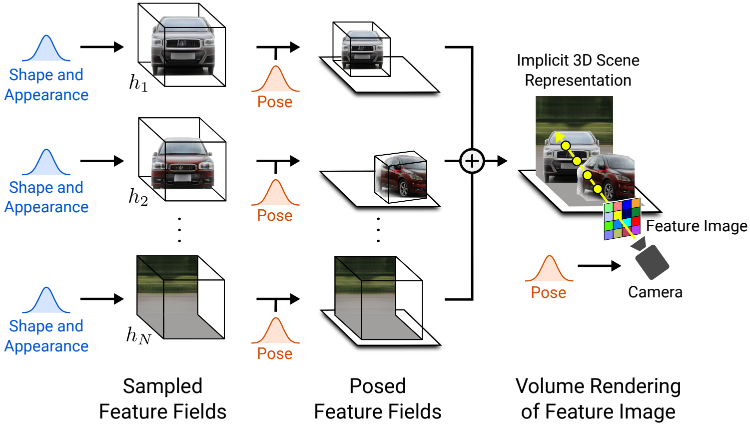
Sample a camera pose.



Niemeyer, Geiger: GIRAFFE: Representing Scenes as Compositional Generative Neural Feature Fields. CVPR, 2021

GIRAFFE

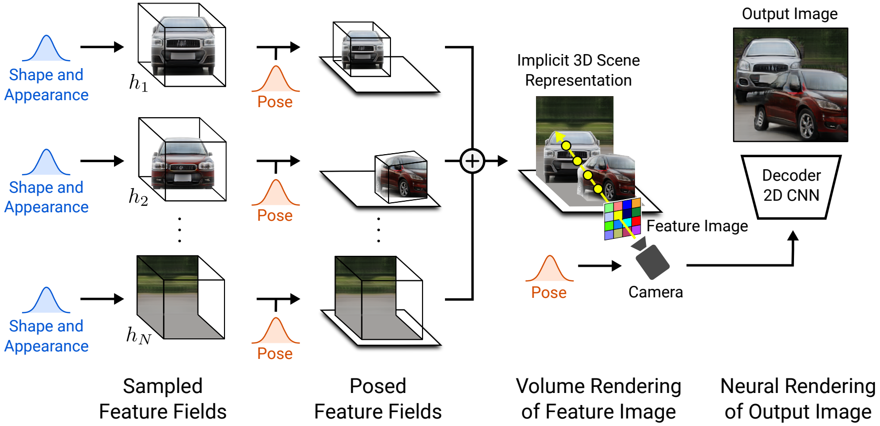
Perform volume rendering and get feature image.



Niemeyer, Geiger: GIRAFFE: Representing Scenes as Compositional Generative Neural Feature Fields. CVPR, 2021

GIRAFFE

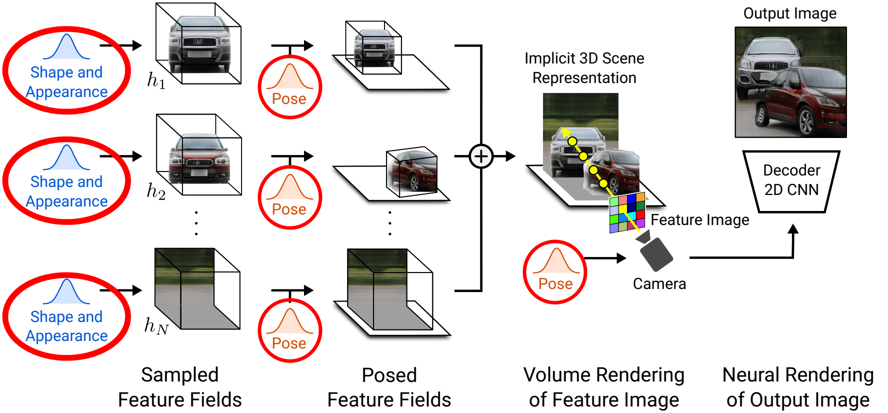
Pass feature image to neural renderer to obtain final output.



Niemeyer, Geiger: GIRAFFE: Representing Scenes as Compositional Generative Neural Feature Fields. CVPR, 2021

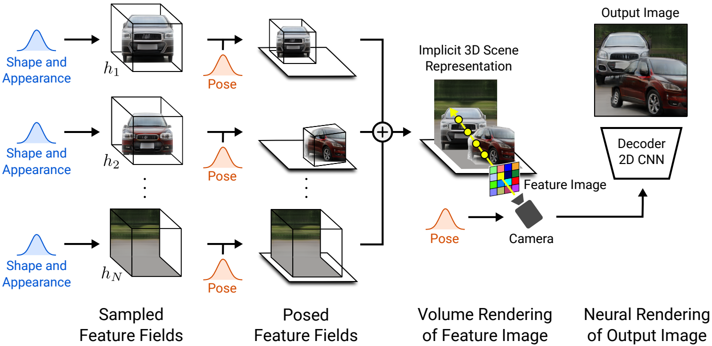
GIRAFFE

At test time, we can sample individual codes and **control the poses**.



Niemeyer, Geiger: GIRAFFE: Representing Scenes as Compositional Generative Neural Feature Fields. CVPR, 2021

GIRAFFE



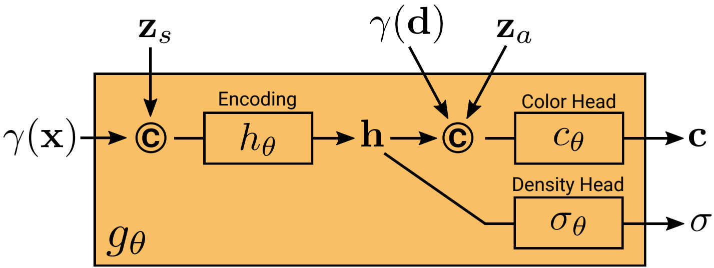
- We train with adversarial loss **on full image**
- We volume-render the feature image at 16×16 pixels

Niemeyer, Geiger: GIRAFFE: Representing Scenes as Compositional Generative Neural Feature Fields. CVPR, 2021

How do we parametrize Feature Fields?

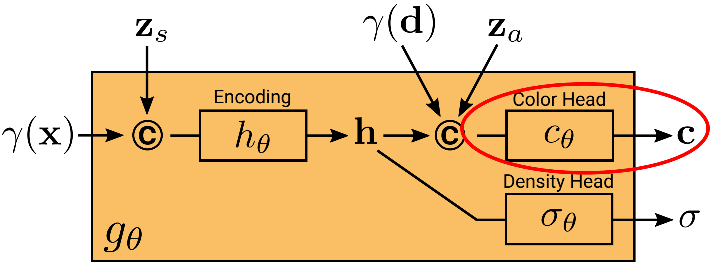
GIRAFFE

Recall the conditional radiance field from before:



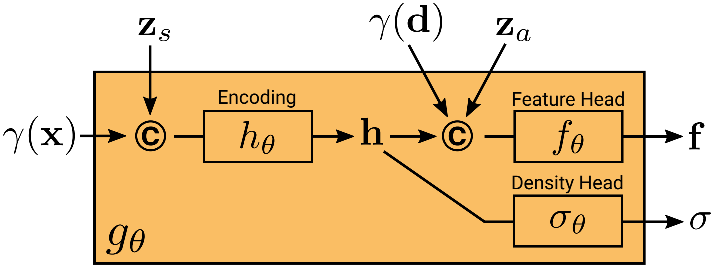
GIRAFFE

We replace the RGB color head with a **feature head**:



GIRAFFE

We replace the RGB color head with a **feature head**:



How do we combine multiple Feature Fields?

GIRAFFE

Scene Composition

We have N feature fields

$$h_i(\mathbf{x}, \mathbf{d}) = (\sigma_i, \mathbf{f}_i)$$

which predict a density σ_i and a feature vector $\mathbf{f}_i$ at $(\mathbf{x}, \mathbf{d})$.

Final density at $(\mathbf{x}, \mathbf{d})$:

$$\sigma = \sum_{i=1}^{N} \sigma_i$$

Final feature vector at $(\mathbf{x}, \mathbf{d})$:

$$\mathbf{f} = \frac{1}{\sigma} \sum_{i=1}^{N} \sigma_i \mathbf{f}_i$$

Niemeyer, Geiger: GIRAFFE: Representing Scenes as Compositional Generative Neural Feature Fields. CVPR, 2021

How well does it work?

GIRAFFE

We compare object translation for a 2D-based GAN (left) and our method (right):

GIRAFFE

We can perform more complex operations like circular translations

Niemeyer, Geiger: GIRAFFE: Representing Scenes as Compositional Generative Neural Feature Fields. CVPR, 2021

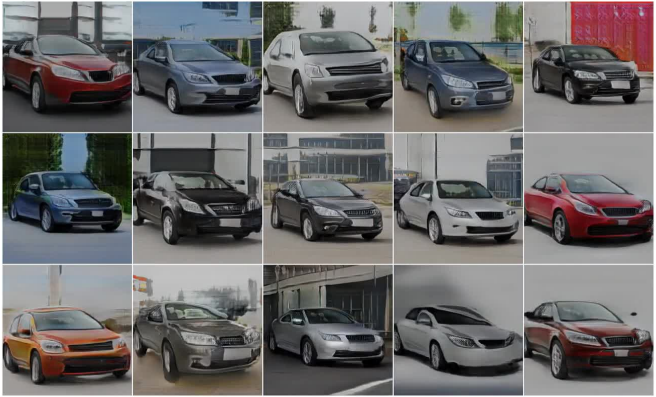
GIRAFFE

We can add more objects at test time (trained on two-object)

Niemeyer, Geiger: GIRAFFE: Representing Scenes as Compositional Generative Neural Feature Fields. CVPR, 2021

GIRAFFE

We can rotate the object



Niemeyer, Geiger: GIRAFFE: Representing Scenes as Compositional Generative Neural Feature Fields. CVPR, 2021

GIRAFFE

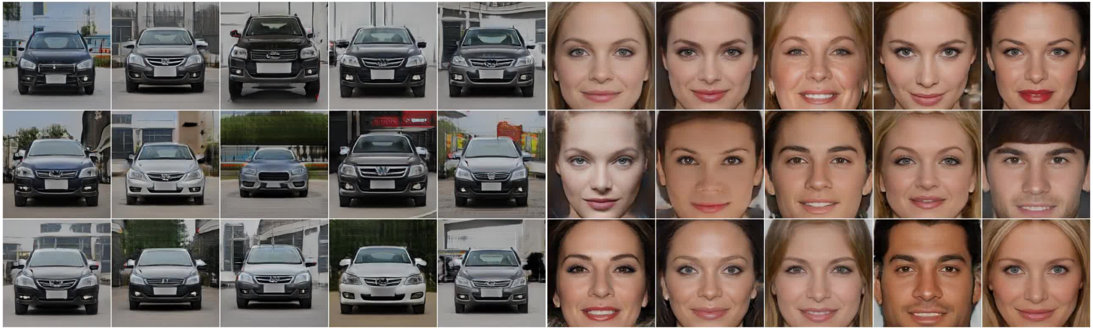
We can translate the object

GIRAFFE

We can change the object shape

GIRAFFE

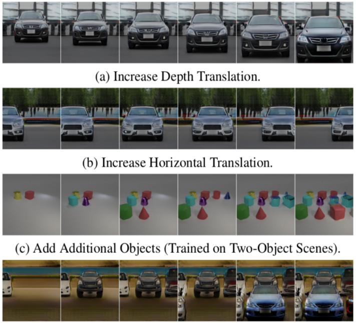
We can change the object appearance



Niemeyer, Geiger: GIRAFFE: Representing Scenes as Compositional Generative Neural Feature Fields. CVPR, 2021

GIRAFFE

We can generate out-of-distribution samples



(a) Increase Depth Translation.

(b) Increase Horizontal Translation.

(c) Add Additional Objects (Trained on Two-Object Scenes).

(d) Add Additional Objects (Trained on Single-Object Scenes).

Total Rendering Time

	64×64	256×256
GRAF	110.1ms	1595.0ms
GIRAFFE	4.8ms	5.9ms

- CNN-based neural renderer yields faster inference.
- We always volume-render the feature image at 16×16 pixels.

How can we scale to more complex camera distributions?

CAMPARI

GRAF, GIRAFFE:

- Learn a 3D-aware image generator with uniform prior on camera distributions
- Requires careful tuning and results degrade if they do not match the data distribution

Niemeyer, Geiger: CAMPARI: Camera-Aware Decomposed Generative Neural Radiance Fields. 3DV, 2021

CAMPARI

GRAF, GIRAFFE:

- Learn a 3D-aware image generator with uniform prior on camera distributions
- Requires careful tuning and results degrade if they do not match the data distribution

CAMPARI:

- Learn a 3D aware image generator and a **camera generator** jointly.

Niemeyer, Geiger: CAMPARI: Camera-Aware Decomposed Generative Neural Radiance Fields. 3DV, 2021

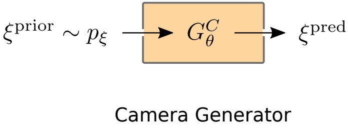
CAMPARI

Sample prior camera $\boldsymbol{\xi}^{\text{prior}} \sim p_{\xi}$.

$\xi^{\text{prior}} \sim p_{\xi}$

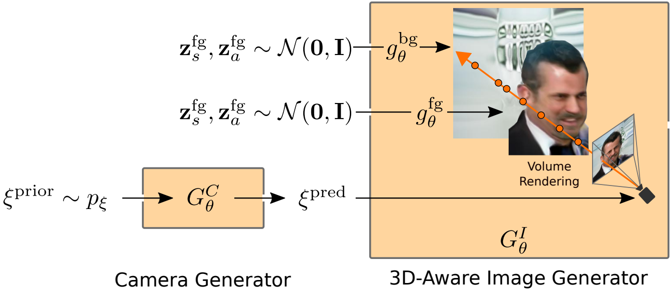
CAMPARI

Pass $\boldsymbol{\xi}^{\text{prior}}$ to camera generator G_θ^C and obtain predicted camera ξ^{pred}.



CAMPARI

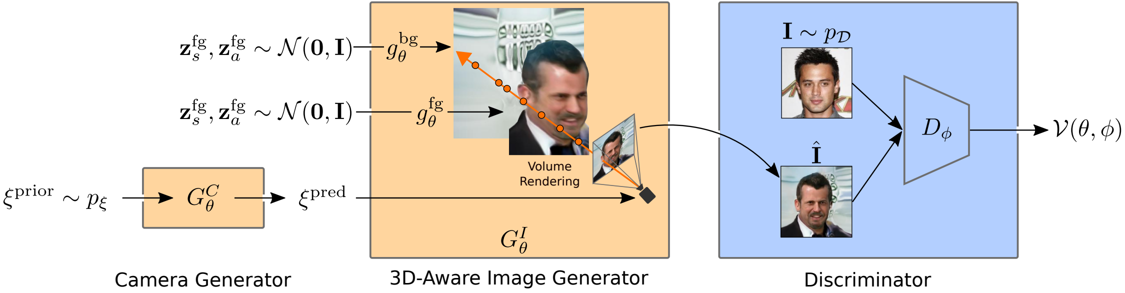
Pass ξ^{pred} and sampled FG / BG latent codes to 3D-aware image generator



Niemeyer, Geiger: CAMPARI: Camera-Aware Decomposed Generative Neural Radiance Fields. 3DV, 2021

CAMPARI

Train entire method with GAN objective (similar to GRAF, GIRAFFE)

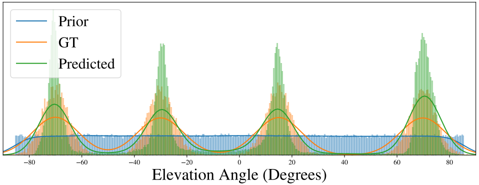


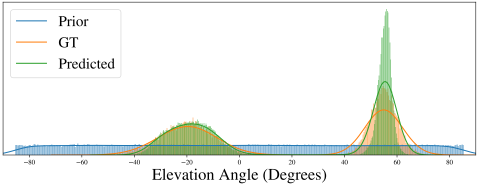
Niemeyer, Geiger: CAMPARI: Camera-Aware Decomposed Generative Neural Radiance Fields. 3DV, 2021

How well does it work?

CAMPARI

CAMPARI learns to match the GT distribution for synthetic datasets





Niemeyer, Geiger: CAMPARI: Camera-Aware Decomposed Generative Neural Radiance Fields. 3DV, 2021

CAMPARI

Results on CelebA

(a) Rotation for GRAF [58] (Camera Parameters Tuned)

(b) Rotation for GRAF [58] (Camera Parameters not Tuned)

(c) Rotation for Ours (No Tuning Required)

Niemeyer, Geiger: CAMPARI: Camera-Aware Decomposed Generative Neural Radiance Fields. 3DV, 2021

Generative Neural Scene Representations

Summary

Generative Neural Scene Representations

Summary

- We propose novel methods for 3D controllable image synthesis

Generative Neural Scene Representations

Summary

- We propose novel methods for 3D controllable image synthesis
- Train from **raw, unposed image collections**

Generative Neural Scene Representations

Summary

- We propose novel methods for 3D controllable image synthesis
- Train from **raw, unposed image collections**
- We incorporate **compositional 3D scene structure** into the generative model

Generative Neural Scene Representations

Summary

- We propose novel methods for 3D controllable image synthesis
- Train from **raw, unposed image collections**
- We incorporate **compositional 3D scene structure** into the generative model
- We have explicit control over **individual objects** during synthesis

Generative Neural Scene Representations

Summary

- We propose novel methods for 3D controllable image synthesis
- Train from **raw, unposed image collections**
- We incorporate **compositional 3D scene structure** into the generative model
- We have explicit control over **individual objects** during synthesis
- Future research: scale to more complex multi-object scenes

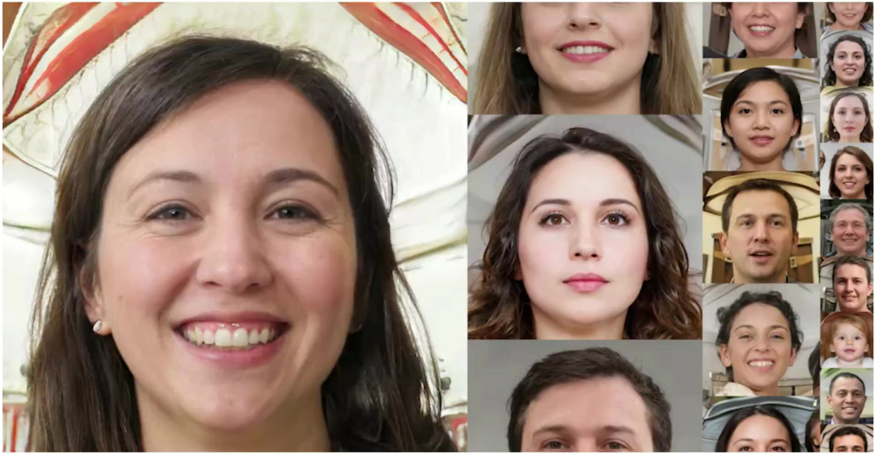
Generative Neural Scene Representations

Summary

- We propose novel methods for 3D controllable image synthesis
- Train from **raw, unposed image collections**
- We incorporate **compositional 3D scene structure** into the generative model
- We have explicit control over **individual objects** during synthesis
- Future research: scale to more complex multi-object scenes
- Future research: disentangle lighting, materials, etc.

Summary

This research is very activate and leads to state-of-the-art results:



Gu, Jiatao, et al. ”StyleNeRF: A Style-based 3D-Aware Generator for High-resolution Image Synthesis.” arXiv preprint arXiv:2110.08985 (2021).

Thank you!

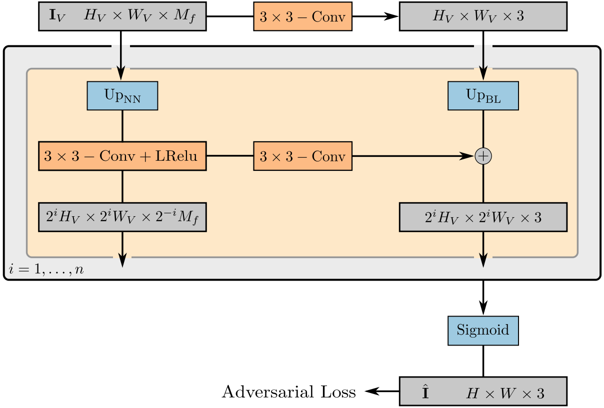
For more information, check out

https://m-niemeyer.github.io/

Appendix

Appendix

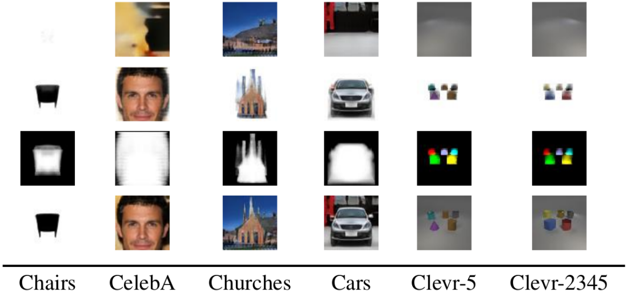
Neural Renderer Architecture



Gu, Jiatao, et al. "StyleNeRF: A Style-based 3D-Aware Generator for High-resolution Image Synthesis." arXiv preprint arXiv:2110.08985 (2021).

Appendix

Disentanglement Results



Gu, Jiatao, et al. "StyleNeRF: A Style-based 3D-Aware Generator for High-resolution Image Synthesis." arXiv preprint arXiv:2110.08985 (2021).

Appendix

Quantitative Results

	Chairs	Cats	CelebA	Cars	Churches
2D GAN [57]	59	18	15	**16**	19
Plat. GAN [31]	199	318	321	299	242
HoloGAN [62]	59	27	25	17	31
GRAF [76]	34	26	25	39	38
Ours	**20**	**8**	**6**	**16**	**17**

Table 1: **Quantitative Comparison.** We report the FID score (↓) at 64^2 pixels for baselines and our method.

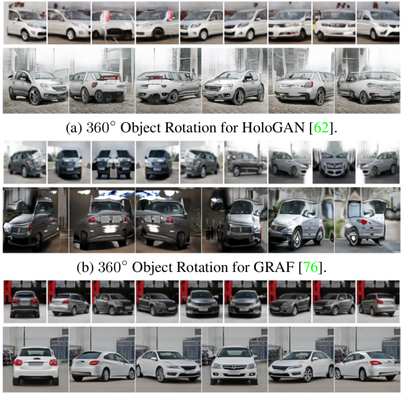
	CelebA-HQ	FFHQ	Cars	Churches	Clevr-2
HoloGAN [62]	61	192	34	58	241
w/o 3D Conv	33	70	49	66	273
GRAF [76]	49	59	95	87	106
Ours	**21**	**32**	**26**	**30**	**31**

Table 2: **Quantitative Comparison.** We report the FID score (↓) at 256^2 pixels for the strongest 3D-aware baselines and our method.

Gu, Jiatao, et al. "StyleNeRF: A Style-based 3D-Aware Generator for High-resolution Image Synthesis." arXiv preprint arXiv:2110.08985 (2021).

Appendix

Baseline Comparison



(a) 360° Object Rotation for HoloGAN [62].

(b) 360° Object Rotation for GRAF [76].

(c) 360° Object Rotation for Our Method.

Gu, Jiatao, et al. ”StyleNeRF: A Style-based 3D-Aware Generator for High-resolution Image Synthesis.” arXiv preprint arXiv:2110.08985 (2021).