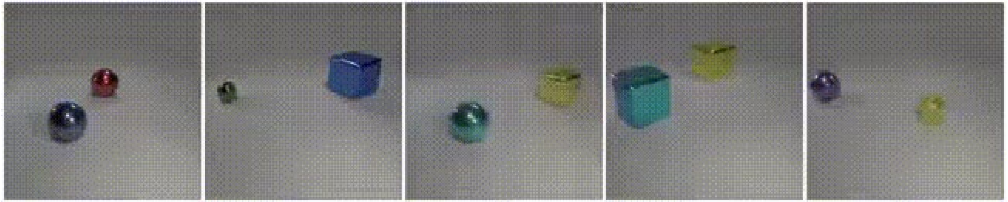
GIRAFFE: Representing Scenes as Compositional Generative Neural Feature Fields

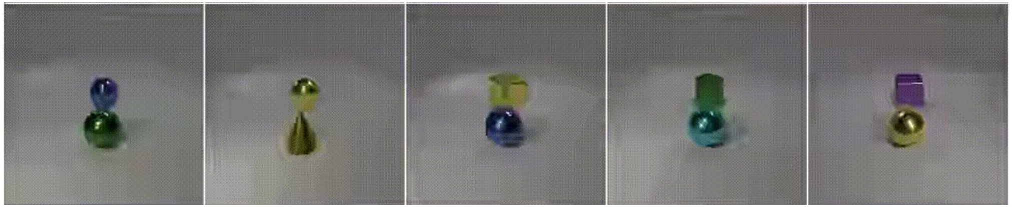
Michael Neimeyer, Andreas Geiger

Presented by **Haofeng Chen**

Controllable Image Generation needs disentanglement



2D CNN

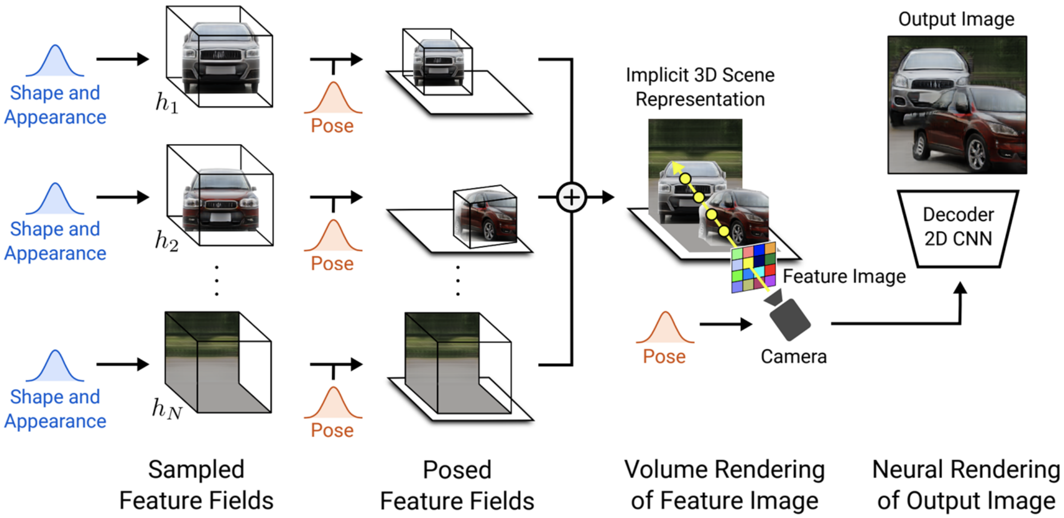


GIRAFFE

- Controlling a single object in the image should not change irrelevant objects
- Disentanglement is hard in 2D generative models

Niemeyer, Michael, and Andreas Geiger. "Giraffe: Representing scenes as compositional generative neural feature fields." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021.

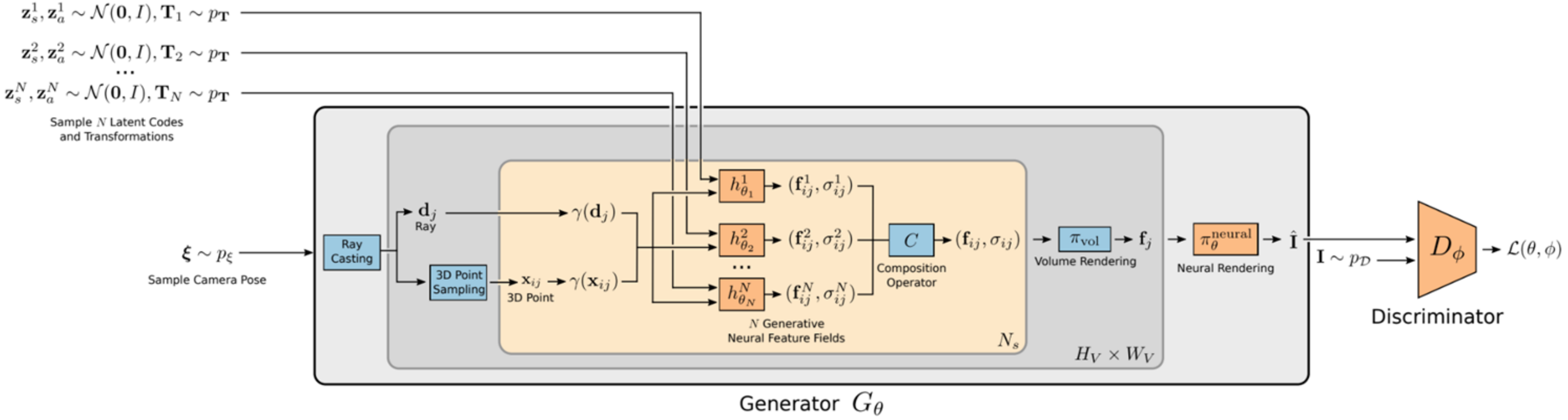
GIRAFFE: Construction



- Represent a scene as compositional generative neural feature fields
- (N - 1) foreground feature fields
- One background feature field

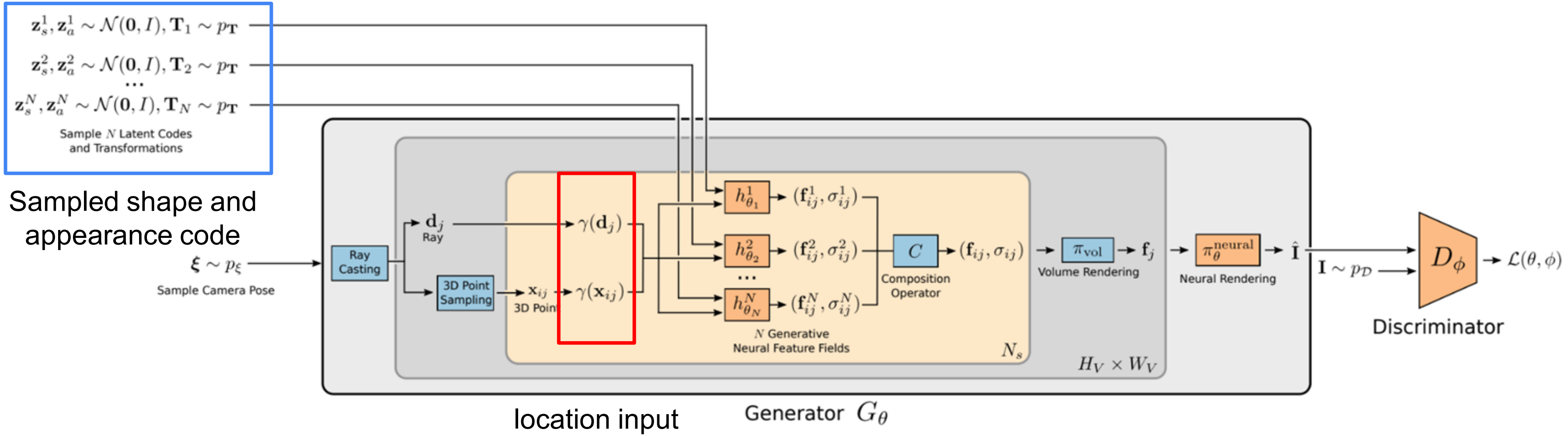
Niemeyer, Michael, and Andreas Geiger. "Giraffe: Representing scenes as compositional generative neural feature fields." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021.

GIRAFFE Architecture



Niemeyer, Michael, and Andreas Geiger. "Giraffe: Representing scenes as compositional generative neural feature fields." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021.

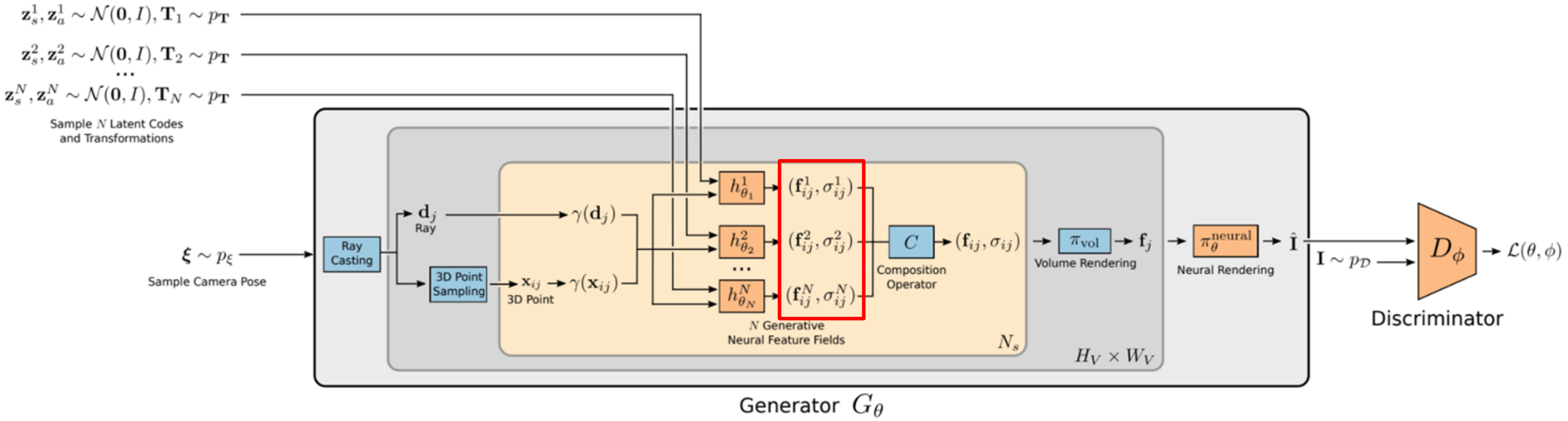
Generative Neural Feature Fields from GRAF



$$
\begin{aligned}
h_\theta : \mathbb{R}^{L_{\mathbf{x}}} \times \mathbb{R}^{L_{\mathbf{d}}} \times \mathbb{R}^{M_s} \times \mathbb{R}^{M_a} &\to \mathbb{R}^+ \times \mathbb{R}^{M_f} \\
(\gamma(\mathbf{x}), \gamma(\mathbf{d}), \mathbf{z}_s, \mathbf{z}_a) &\mapsto (\sigma, \mathbf{f})
\end{aligned}
\quad (4)
$$

Niemeyer, Michael, and Andreas Geiger. "Giraffe: Representing scenes as compositional generative neural feature fields." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021.

Generative Neural Feature Fields from GRAF

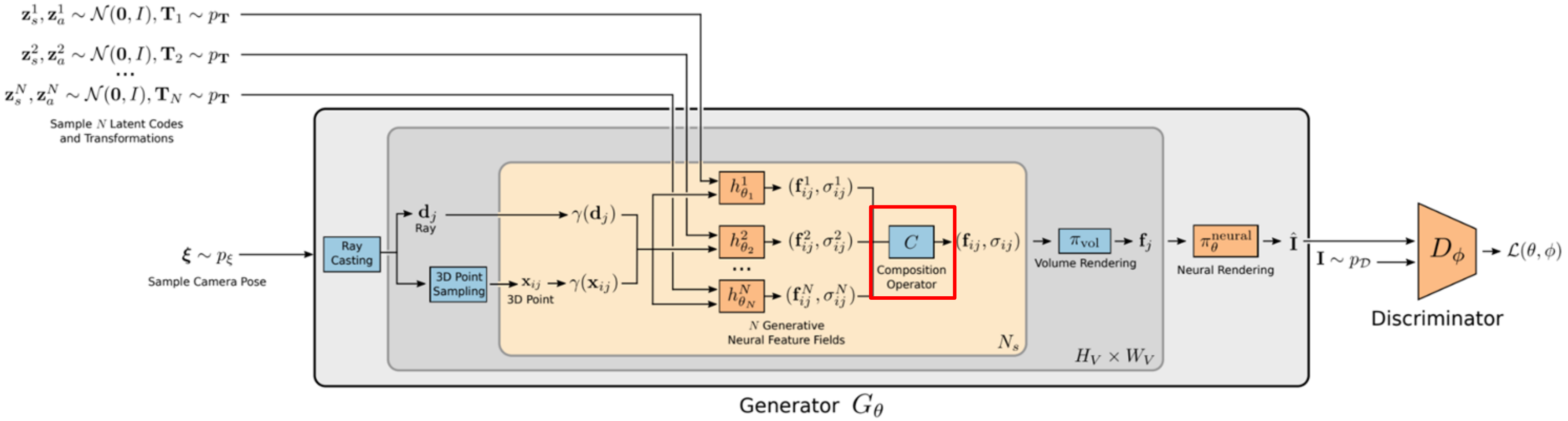


$$h_\theta : \mathbb{R}^{L_\mathbf{x}} \times \mathbb{R}^{L_\mathbf{d}} \times \mathbb{R}^{M_s} \times \mathbb{R}^{M_a} \to \mathbb{R}^+ \times \mathbb{R}^{M_f}$$
$$(\gamma(\mathbf{x}), \gamma(\mathbf{d}), \mathbf{z}_s, \mathbf{z}_a) \mapsto (\sigma, \mathbf{f}) \quad (4)$$

Outputs features instead of colors

Niemeyer, Michael, and Andreas Geiger. "Giraffe: Representing scenes as compositional generative neural feature fields." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021.

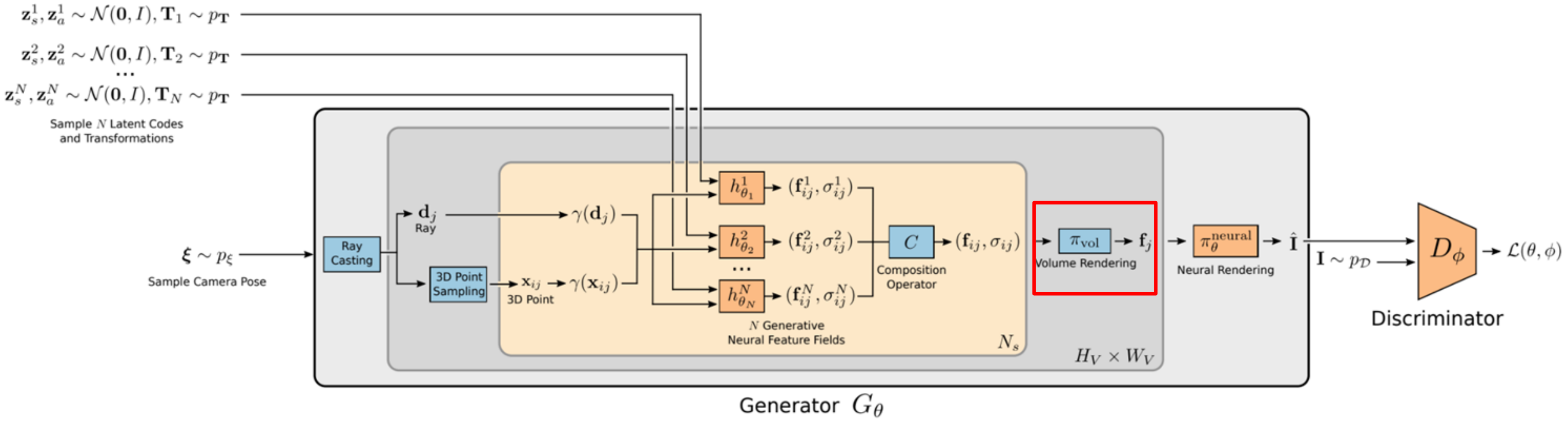
Composition of objects by summation



$$C(\mathbf{x}, \mathbf{d}) = \left(\sigma, \frac{1}{\sigma} \sum_{i=1}^{N} \sigma_i \mathbf{f}_i\right), \text{ where } \quad \sigma = \sum_{i=1}^{N} \sigma_i \quad (8)$$

Niemeyer, Michael, and Andreas Geiger. "Giraffe: Representing scenes as compositional generative neural feature fields." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021.

3D Volume Rendering in feature space

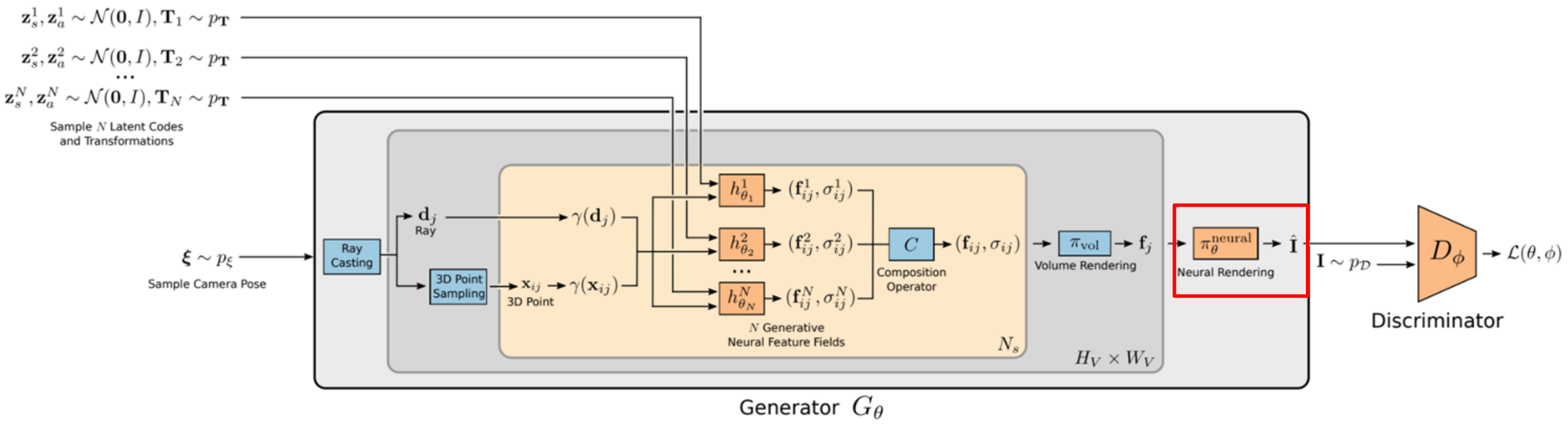


$$\{\sigma_j, \mathbf{f}_j\}_{j=1}^{N_s} \mapsto \mathbf{f} \quad (9)$$

$$\mathbf{f} = \sum_{j=1}^{N_s} \tau_j \alpha_j \mathbf{f}_j \quad \tau_j = \prod_{k=1}^{j-1} (1 - \alpha_k) \quad \alpha_j = 1 - e^{-\sigma_j \delta_j} \quad (10)$$

Niemeyer, Michael, and Andreas Geiger. "Giraffe: Representing scenes as compositional generative neural feature fields." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021.

2D Neural Rendering



$$\pi_\theta^{\text{neural}} : \mathbb{R}^{H_V \times W_V \times M_f} \rightarrow \mathbb{R}^{H \times W \times 3} \qquad (11)$$

Feature space to image space, achieved by 2D CNN

Niemeyer, Michael, and Andreas Geiger. "Giraffe: Representing scenes as compositional generative neural feature fields." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021.

GIRAFFE is better, lighter, and faster at generation

	Cats	CelebA	Cars	Chairs	Churches
2D GAN [58]	18	15	**16**	59	19
Plat. GAN [32]	318	321	299	199	242
BlockGAN [64]	47	69	41	41	28
HoloGAN [63]	27	25	17	59	31
GRAF [77]	26	25	39	34	38
Ours	**8**	**6**	**16**	**20**	**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 [63]	61	192	34	58	241
w/o 3D Conv	33	70	49	66	273
GRAF [77]	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.

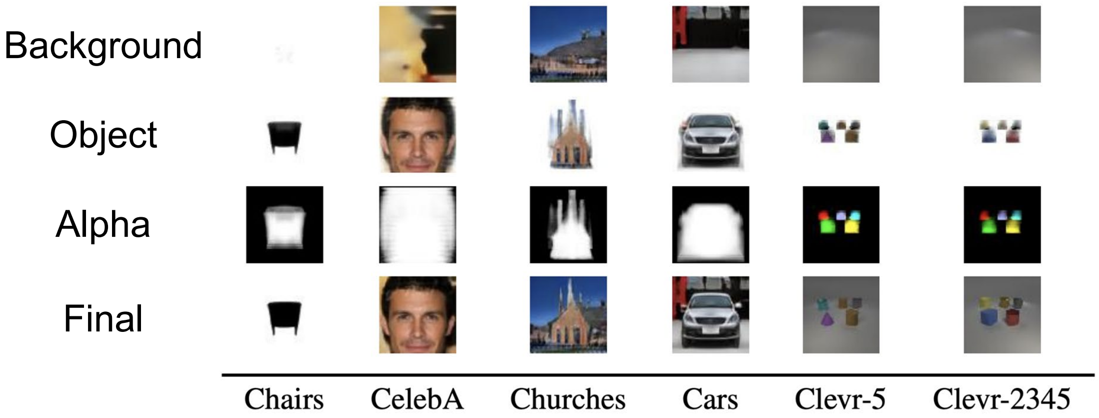
2D GAN	Plat. GAN	BlockGAN	HoloGAN	GRAF	Ours
1.69	381.56	4.44	7.80	0.68	0.41

Table 3: **Network Parameter Comparison.** We report the number of generator network parameters in million.

- Better FID score for all resolutions
- Much less parameters
- Rendering time reduced from 1595.0 ms to 5.9 ms from [77] with 256^2 pixels

Niemeyer, Michael, and Andreas Geiger. "Giraffe: Representing scenes as compositional generative neural feature fields." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021.

Scene disentanglement learned w/o supervision



- Represents foreground and background as separate objects
- Learns to generate background although no complete background is in the dataset (in-painting)

Niemeyer, Michael, and Andreas Geiger. "Giraffe: Representing scenes as compositional generative neural feature fields." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021.

Results: Rotation

Rotation

Niemeyer, Michael, and Andreas Geiger. "Giraffe: Representing scenes as compositional generative neural feature fields." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021.

Results: Translation

Horizontal Translations

Vertical Translations

Niemeyer, Michael, and Andreas Geiger. "Giraffe: Representing scenes as compositional generative neural feature fields." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021.

Results: Changing Foreground / Background

Changing Object Appearance

Changing Background

Niemeyer, Michael, and Andreas Geiger. "Giraffe: Representing scenes as compositional generative neural feature fields." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021.

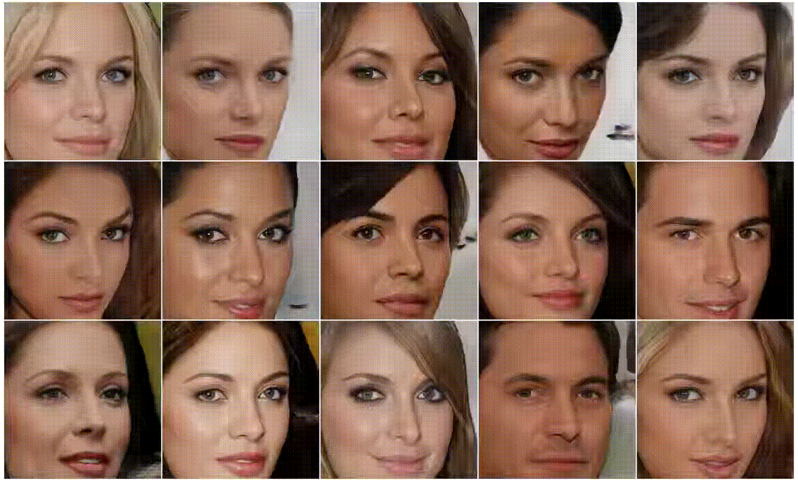
Results: Adding Objects

Adding Objects in CLEVR

Adding Object in Cars (OOD)

Niemeyer, Michael, and Andreas Geiger. "Giraffe: Representing scenes as compositional generative neural feature fields." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021.

Limitation: Bias of Dataset



Rotation in CelebA

Niemeyer, Michael, and Andreas Geiger. "Giraffe: Representing scenes as compositional generative neural feature fields." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021.