GIRAFFE: Representing Scenes as Compositional Generative Neural Feature Fields

Michael Neimeyer, Andreas Geiger

Presented by Haofeng Chen
Controllable Image Generation needs disentanglement

- 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

\[ \begin{align*}
    z_1, z_a &\sim \mathcal{N}(0, I), T_1 \sim p_T \\
    z_2, z_2^a &\sim \mathcal{N}(0, I), T_2 \sim p_T \\
    ... \\
    z_N, z_N^a &\sim \mathcal{N}(0, I), T_N \sim p_T \\
\end{align*} \]

Sample \( \mathcal{N} \) Latent Codes and Transformations

\[ \gamma(d_j) \]

Sampled shape and appearance code

Location input

Generator \( G_\theta \)

\[ h_\theta : \mathbb{R}^{L_x} \times \mathbb{R}^{L_d} \times \mathbb{R}^{M_s} \times \mathbb{R}^{M_a} \rightarrow \mathbb{R}^+ \times \mathbb{R}^{M_f} \]

\[ (\gamma(x), \gamma(d), z_s, z_a) \mapsto (\sigma, f) \]  

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

\[ z^1_s, z^1_a \sim N(0, I), T_1 \sim p_T \]
\[ z^2_s, z^2_a \sim N(0, I), T_2 \sim p_T \]
\[ \vdots \]
\[ z^N_s, z^N_a \sim N(0, I), T_N \sim p_T \]

Sample \( N \) Latent Codes and Transformations

\[ \xi \sim p_{\xi} \]
Sample Camera Pose

\[ \rho \rightarrow \mathcal{C}(\theta, \phi) \]

Discriminator

\[ h_{\theta} : \mathbb{R}^{L_x} \times \mathbb{R}^{L_d} \times \mathbb{R}^{M_s} \times \mathbb{R}^{M_a} \rightarrow \mathbb{R}^+ \times \mathbb{R}^{M_f} \]

(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

\[ C(x, d) = \left( \sigma - \frac{1}{\sigma} \sum_{i=1}^{N} \sigma_i f_i \right), \text{ where } \sigma = \sum_{i=1}^{N} \sigma_i \]  

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

\[ \{\sigma_j, f_j\}_{j=1}^{N_a} \mapsto f \quad (9) \]

\[ f = \sum_{j=1}^{N_a} \tau_j \alpha_j f_j \quad \tau_j = \prod_{k=1}^{j-1} (1 - \alpha_k) \quad \alpha_j = 1 - e^{-\sigma_j \delta_j} \quad (10) \]

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

\[ \pi^\text{neural}_\theta : \mathbb{R}^{H_V \times W_V \times M_f} \rightarrow \mathbb{R}^{H \times W \times 3} \] (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 64² pixels for baselines and our method.

|                      | CelebA-HQ | FFHQ | Cars | Churches | CleVR-2 |
|----------------------|------------|------|------|----------|---------|
| HoloGAN [63] w/o 3D Conv | 61  | 192  | 34   | 58       | 241     |
|                      | 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 256² 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 256² 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

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

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

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

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.