JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Polynomial Implicit Neural Representations For Large Diverse Datasets

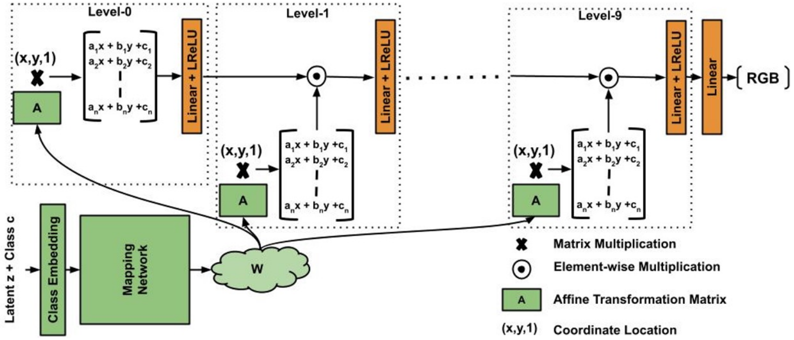
Rajhans Singh, Ankita Shukla, Pavan Turaga

Geometric Media Lab, Arizona State University

Arizona State University

Overview

- **Implicit Neural Representations** (INRs) are effective for 2D or 3D scene representation and are further extended as a generative model.
- **Sinusoidal position encoding** : positional embedding can limit large dataset representation.
- We propose **Polynomial Implicit Neural Representation (Poly-INR)** and design an MLP model for the approximation of higher-order polynomials.
- Poly-INR as a generative model performs comparably to the state-of-the-art CNN-based generative models on the ImageNet dataset.



Implicit Neural Representation (INR)

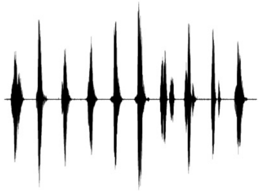
- **Classical signal representation**: discretize pixel value, point cloud, discretize amplitude.

- **Implicit neural representation** (**INR**): Multi-Layer Perceptron (MLP) is trained to generate the signal.

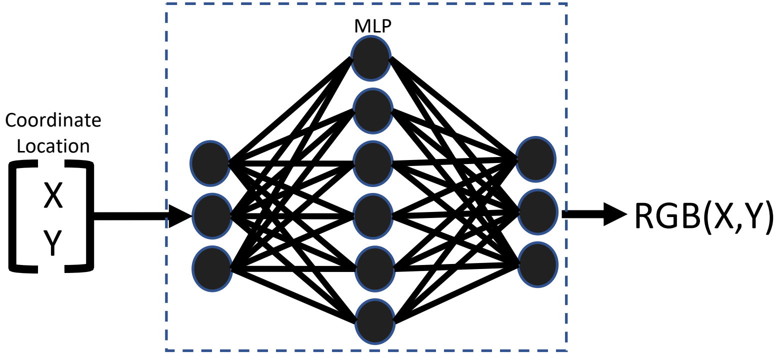
- **Benefits of such representation:**
    - Memory efficient
    - Better gradients or higher order derivatives computation
    - Solving inverse problem

Image

3D Shape

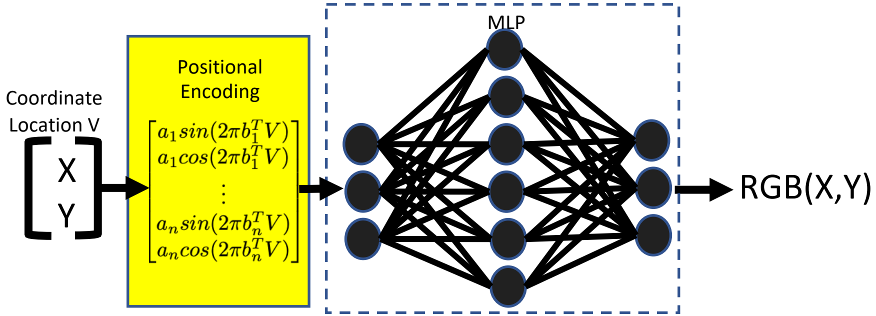


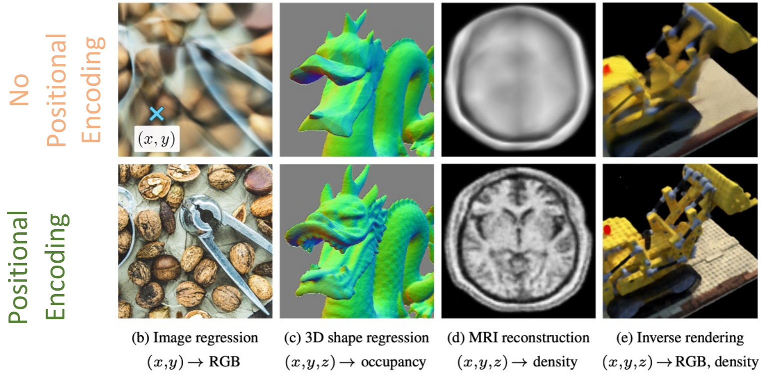
Audio



Sitzmann et al. “Implicit neural representations with periodic activation functions”, Advances in Neural Information Processing Systems (2020).

Implicit Neural Representation (INR)





- ReLU-based MLP only retains low-frequency information.
- Periodic function based positional encoding is used for high frequency representation.
- **Positional encoding is limiting** for large dataset representation for two reasons:
  - Size of the encoding space is limited and low dimensional.
  - Conventional CNN based generative model first represents low frequency information like shape in the initial levels and progressively adds high frequency information.

Tancik et al. "Fourier features let networks learn high frequency functions in low dimensional domains", Advances in Neural Information Processing Systems (2020).

Polynomial Implicit Neural Representation

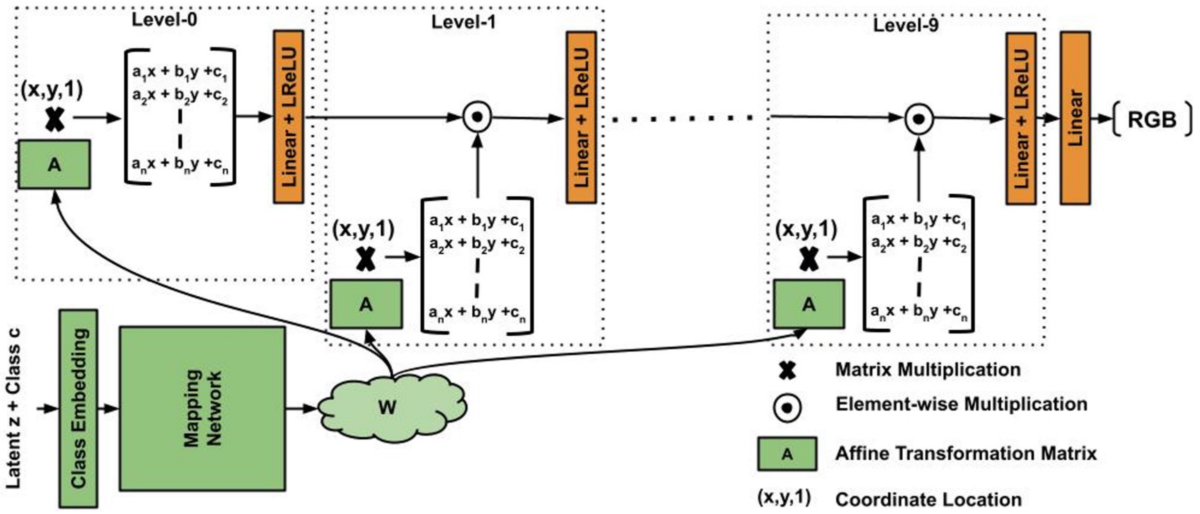
- Polynomial representation:

$$G(x, y) = g_{00} + g_{10}x + g_{01}y + \ldots + g_{pq}x^p y^q$$

where *(x, y)* is the normalized pixel location and polynomial coefficients g_{pq} are parameterized by a latent vector *z*.

- Positional embedding of the form $x^p y^q$ to approximate a higher-order polynomial can be limiting due to finite size of embedding space.
- Hence, we progressively increase the polynomial order in the network and let it learn the required orders.
- We use element-wise multiplication with the affine-transformed coordinate location at different levels, giving the network flexibility to flexibility to increase the order as required.
- Model is trained as a Generative Adversarial Network (GAN).

Polynomial Implicit Neural Representation



- **Mapping network** takes the latent code $z \in R^{64}$ and maps it to the affine parameter space $W \in R^{512}$, consists of consists of an MLP with two linear layers.
- Synthesis network generates the RGB value for given pixel location.
- **Synthesis Network**: $G_{syn} = \ldots \sigma(W_2((A_2X) \odot \sigma(W_1((A_1X) \odot \sigma(W_0(A_0X)))))),$

Quantitative Results

- Comparison against CNN-based GANs (Big- GAN and StyleGAN-XL) and diffusion models (ADM and DiT-XL) on the ImageNet dataset.
- Metrics: Frechet Inception Distance (FID), Inception Score (IS), Precision (Pr) and Recall (Rec).

(a) ImageNet 128 × 128

Model	FID ↓	sFID ↓	rFID ↓	IS ↑	Pr ↑	Rec ↑
BigGAN	6.02	7.18	6.09	145.83	0.86	0.35
CDM	3.52	-	-	128.80	-	-
ADM	5.91	5.09	13.29	93.31	0.70	0.65
ADM-G	2.97	5.09	3.80	141.37	0.78	0.59
StyleGAN-XL	1.81	3.82	1.82	200.55	0.77	0.55
Poly-INR	2.08	3.93	2.76	179.64	0.70	0.45

(b) ImageNet 256 × 256

Model	FID ↓	sFID ↓	rFID ↓	IS ↑	Pr ↑	Rec ↑
BigGAN	6.95	7.36	75.24	202.65	0.87	0.28
ADM	10.94	6.02	125.78	100.98	0.69	0.63
ADM-G	3.94	6.14	11.86	215.84	0.83	0.53
DiT-XL/2-G	2.27	4.60	-	278.54	0.83	0.57
StyleGAN-XL	2.30	4.02	7.06	265.12	0.78	0.53
Poly-INR	2.86	4.37	7.79	241.43	0.71	0.39

(c) ImageNet 512 × 512

Model	FID ↓	sFID ↓	rFID ↓	IS ↑	Pr ↑	Rec ↑
BigGAN	8.43	8.13	312.00	177.90	0.88	0.29
ADM	23.24	10.19	561.32	58.06	0.73	0.60
ADM-G	3.85	5.86	210.83	221.72	0.84	0.53
DiT-XL/2-G	3.04	5.04	-	240.82	0.84	0.54
StyleGAN-XL	2.41	4.06	51.54	267.75	0.77	0.52
Poly-INR	3.81	5.06	54.31	267.44	0.70	0.34

(d) Number of parameters in millions (M)

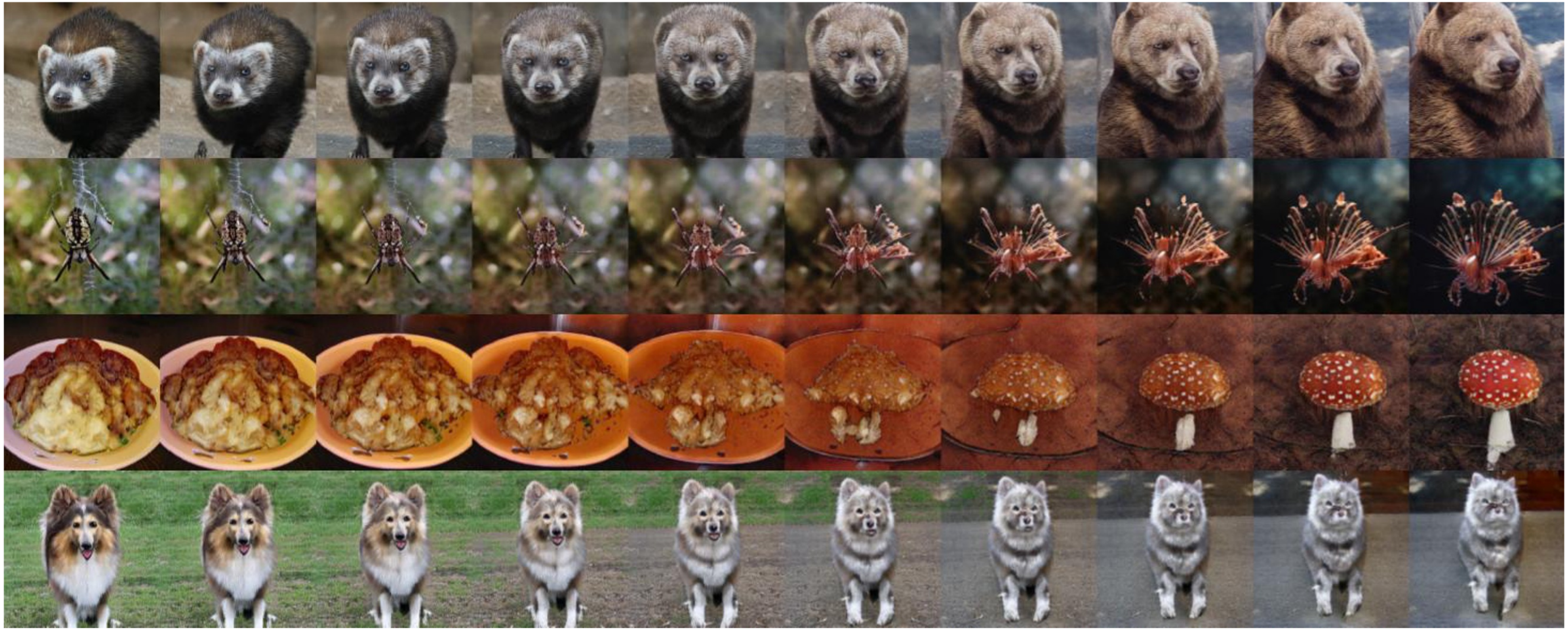
Model	64^2	128^2	256^2	512^2
BigGAN	-	141.0	164.3	164.7
ADM	296.0	422.0	554.0	559.0
DiT-XL	-	-	675.0	675.0
StyleGAN-XL	134.4	158.7	166.3	168.4
Poly-INR	46.0	46.0	46.0	46.0

Sauer, Axel, Katja Schwarz, and Andreas Geiger. "Stylegan-xl: Scaling stylegan to large diverse datasets." ACM SIGGRAPH 2022 conference proceedings. 2022.
Dhariwal, Prafulla, and Alexander Nichol. "Diffusion models beat gans on image synthesis." Advances in Neural Information Processing Systems 34 (2021)
Peebles, William, and Saining Xie. "Scalable Diffusion Models with Transformers.". 2022

Generated Samples

Interpolation

Poly-INR provides smooth interpolation in affine parameters space.

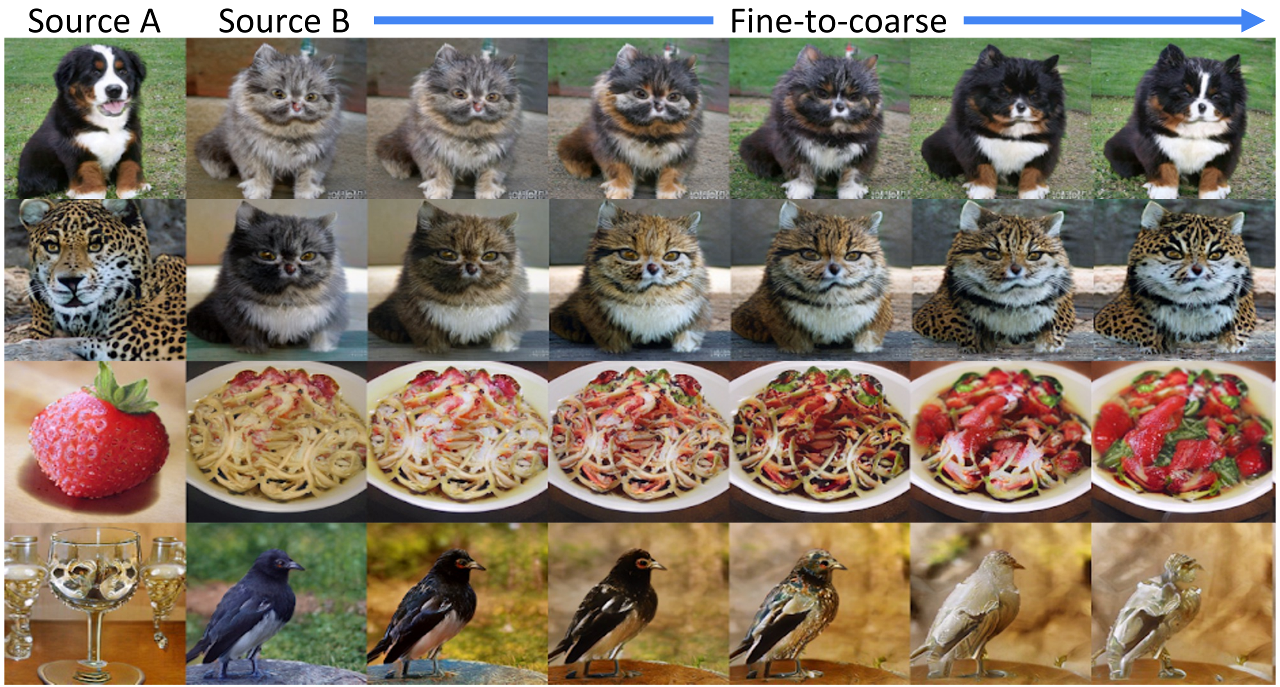


Shape Mixing

Copying the affine parameters of source A to source B at lower levels (0-5) brings change in the shape

Style Mixing

Copying the affine parameters of source A to source B at higher levels (5 - 9) brings a change in the style



Thank You