Novel Shape Generation with SO(3)-Equivariant Auto-Encoders

Ethan Chun [1] and Shayan Pardis [1]

[1]Massachusetts Institute of Technology

Abstract

- **Context.** Automated 3D object generation can revolutionize fields from 3D content creation to robotics.
- **Problem.** Current data representations are inefficient to produce with learning approaches.
- **Contribution.** By parameterizing 3D data as a signal in the spherical harmonic basis, learning an equivariant autoencoder, and showing equivariant latent space interpolation, we provide a foundation for future work in equivariant 3D shape generation.

Data Representation

Spherical Harmonics for 3D Shapes

We represent any star-shaped signal as the coefficients of the spherical harmonics. This format is ideal for equivariant machine learning, increasing data efficiency of potential generative methods.

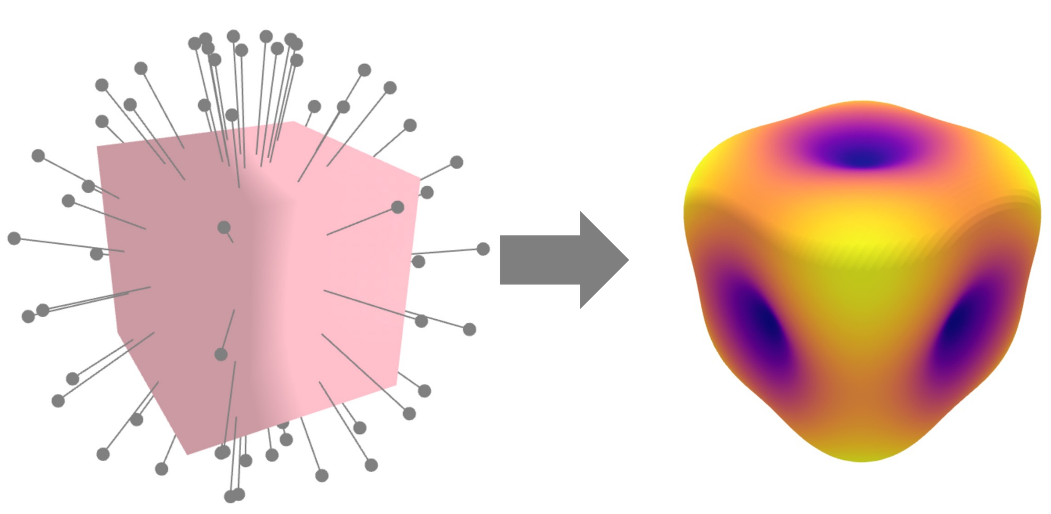

Figure 1. Mesh to Spherical Harmonics – We center the mesh, then cast rays along a α, β grid to get the surface distance from the origin. This grid is used to calculate the spherical harmonic coefficients to degree l_{max}. We set $l_{max} = 4$

Model Architecture

Equivariant Autoencoder

We use an equivariant encoder-decoder architecture derived from Spherical CNNs.

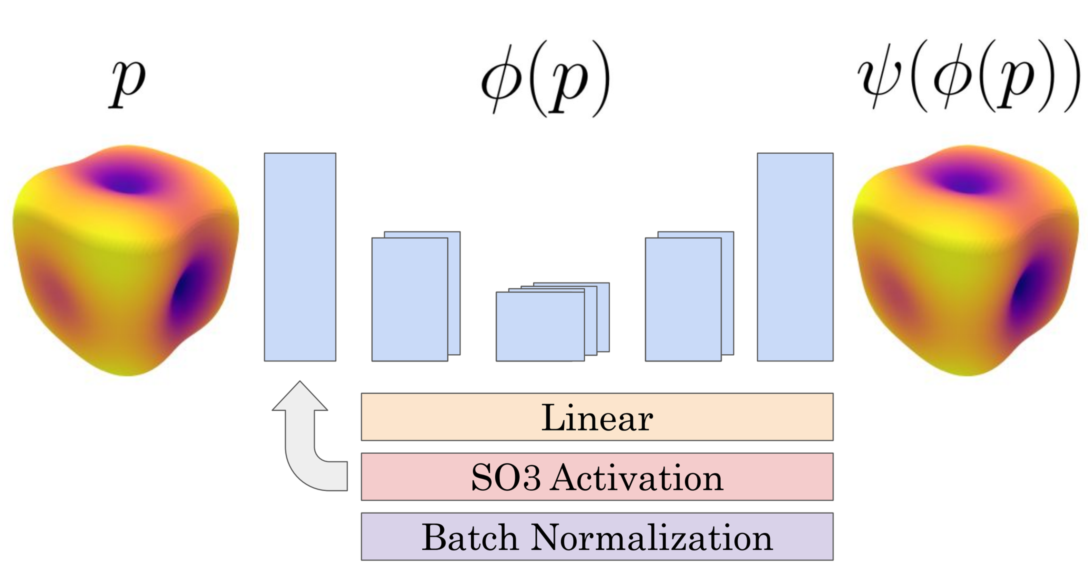

Figure 2. In the encoder, we decrease l_max, the maximum degree of irreps, and increase the number of channels. The latent space only consists of $l = 0, 1$ irreps. The decoder is the mirror of the encoder.

Training and Inference

Loss Functions

MSE loss in Spherical Harmonics Space is not meaningful. Therefore, we sample points on S^2 and compare values of the function at those points.

$$L(f_1, f_2) = \mathbf{E}_{X \sim S^2}\left[\tilde{L}(f_1(X), f_2(X))\right]$$

We used Huber Loss for $\tilde{L}$.

Interpolation

Rotation should be independent of the change of shape. That is, interpolating between p and $g.p$ should only rotate p.

$$\gamma_{p,q}(t) = \psi\left(\left(\frac{\|\phi(q)\|}{\|\phi(p)\|}\right)^t D(g_{\phi(p),\phi(q)})^t \phi(p)\right)$$

In $l > 1$ rotation between vectors $\phi(p), \phi(q)$ is not unique. Thus, latent space should only have $l = 0, 1$

Experiments

Reconstruction

First, we validate our architecture by reconstructing a single 1.0 unit size cube and then reconstructing objects from a dataset of 100 rectangular prisms with side lengths between 0.5 and 2.0 units.

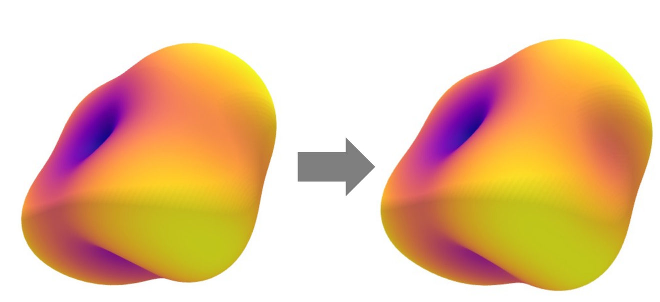

Figure 3. Due to the equivariant design of our network, we demonstrate robust reconstruction of rotated input objects.

Latent Space Interpolation

We explore the generative abilities of our network by testing our equivariant interpolation against a linear latent space traversal.

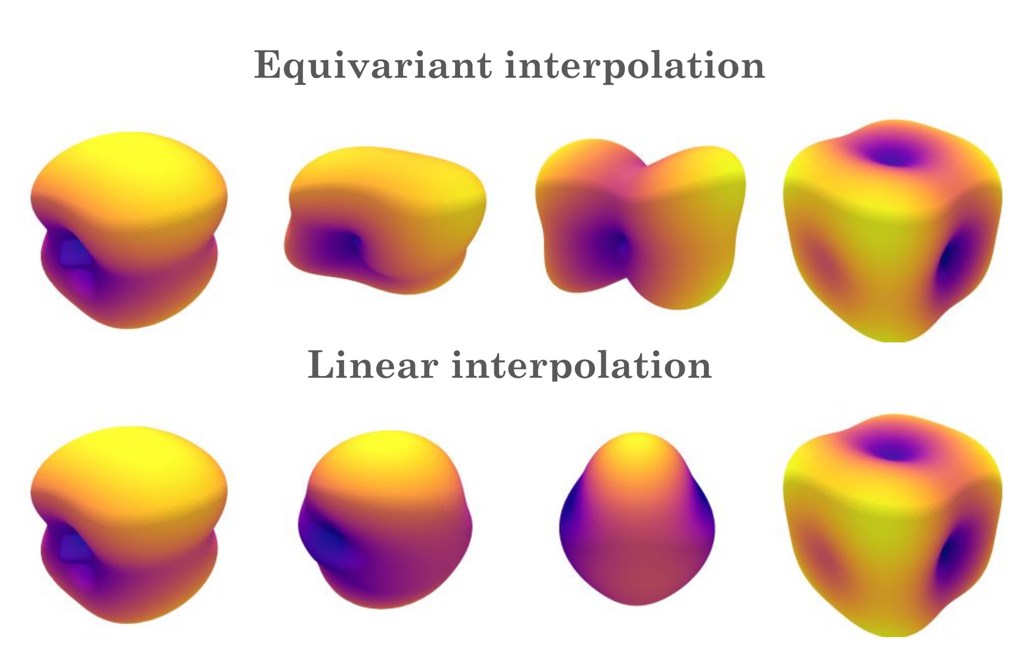

Figure 4. Equivariant interpolation stays in the distribution of dataset (boxes), however, in linear interpolation, out-of-distribution objects appear.

Discussion

Latent Space Dimensionality

The dimensionality of the latent space greatly affects training dynamics.

- With $l_{max} = 2$, training is fast and robust.
- With $l_{max} = 1$, training often fails or converges to predicting a cylinder.

Increasing the depth of the network appeared to improve performance.

Using an equivariant LayerNorm (other than Batch Normalization) will potentially improve the generalization.

Sampling Methods

Our model is approximately equivariant.

- We attribute this to numerical issues when applying activation functions on the SO(3) grid.
- Future work may examine more uniform sampling as a replacement for grid based approaches.

Stable Learning and Normalization

Our model requires Batch Normalization to train. However, state-of-the-art models do not. Future work may investigate applying the LayerNorm techniques in models such as Equiformer.

Alternative Architectures

We also tested tensor products to downsample and upsample, however, training was slow and sensitive to loss. Future work may investigate the cause.

Having a fixed l_{max} is also a fundamental limitation of using spherical harmonics that prevents processing data with arbitrary precision. Tensor-Field eliminates the need for evaluating our function on a grid.

Conclusion

We show promising novel shape generation with equivariant latent space traversals in our equivariant autoencoder.

We introduce the use of spherical harmonics as a data representation for 3D object generation.

Future work may explore more advanced generative equivariant models such as Equivariant Latent Diffusion Models.