SO(3) Diffusion models for Synthetic Galaxy Generation

Yesukhei Jagvaral (PhD student)

Carnegie Mellon University

Advised by:
Rachel Mandelbaum (CMU)
Francois Lanusse (CEA Paris-Saclay/CCA Flatiron)

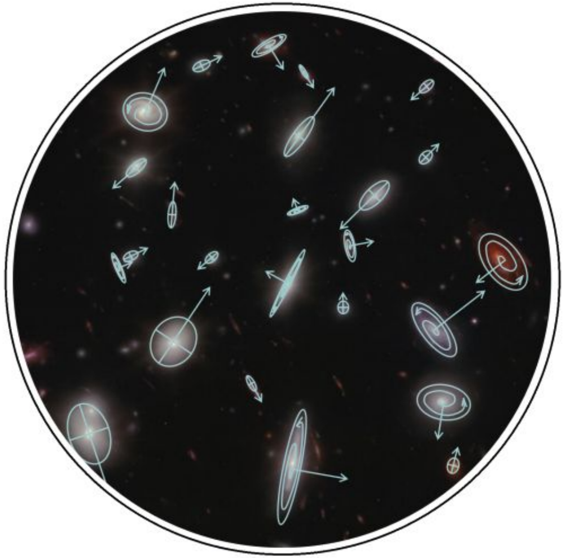
Overview

Problem Statement: Modeling correlated galaxy orientations, so far no single model can describe this on all scales.

Proposed Method: Geometric Deep Learning (Manifolds, Groups), specifically **Diffusion on SO(3)**

Goal: Synthetic **galaxy** catalogs with realistically complex orientations

Results: Our model is able to learn and generate galaxy orientations that are statistically consistent with the reference hydro-simulation

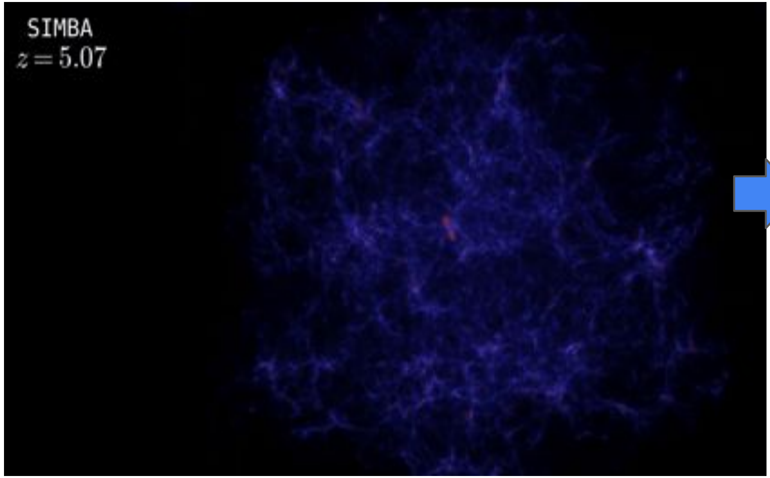


Credit: Lamman et al, 2023

Generating Galaxy Catalogs with Deep Learning

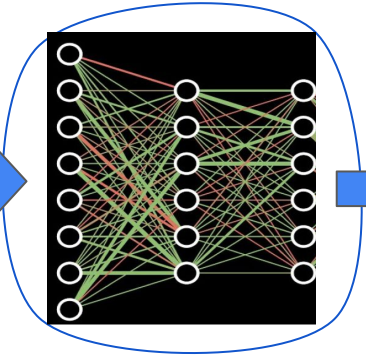
as demanded by large sky surveys with high resolution and volume

Gravity+Galaxy



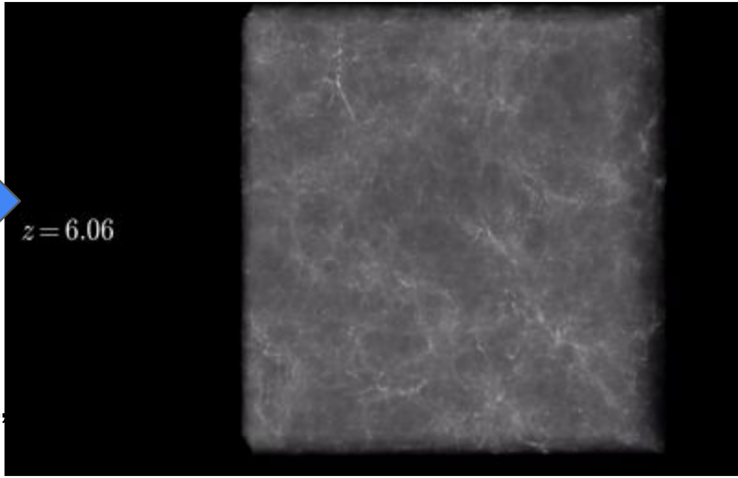
- Volume: Small
- Resolution: High
- Cost: High

Deep Learning



Learn the "**Galaxies**"

Gravity-only simulation

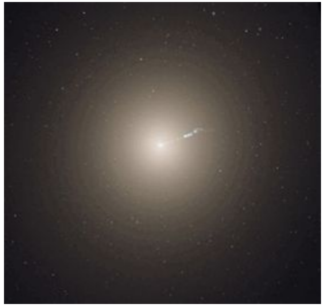


- Volume: Large
- Resolution: Low
- Cost: Low

Credit: 3Blue1Brown, camel-simulations.org

Modeling Galaxy Properties

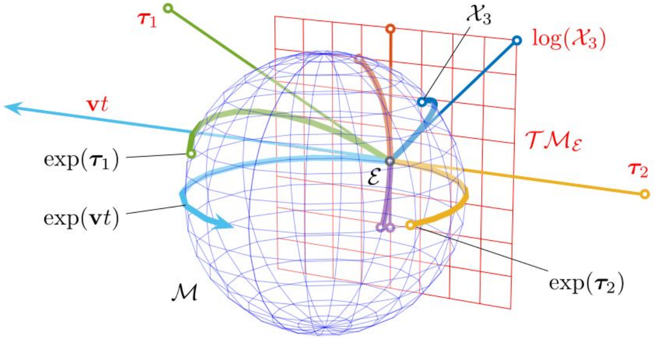
Galaxy orientations in **3D**



Scalars

non-Euclidean manifold

Euclidean manifold



SO(3) - Special orthogonal group of 3D

- Constrained to the 4D hypersphere (Quaternion representation of rotations)

Credit: HST, forbes.com

Galaxy Orientations & Intrinsic Alignment

Dark Energy equation of state parameters

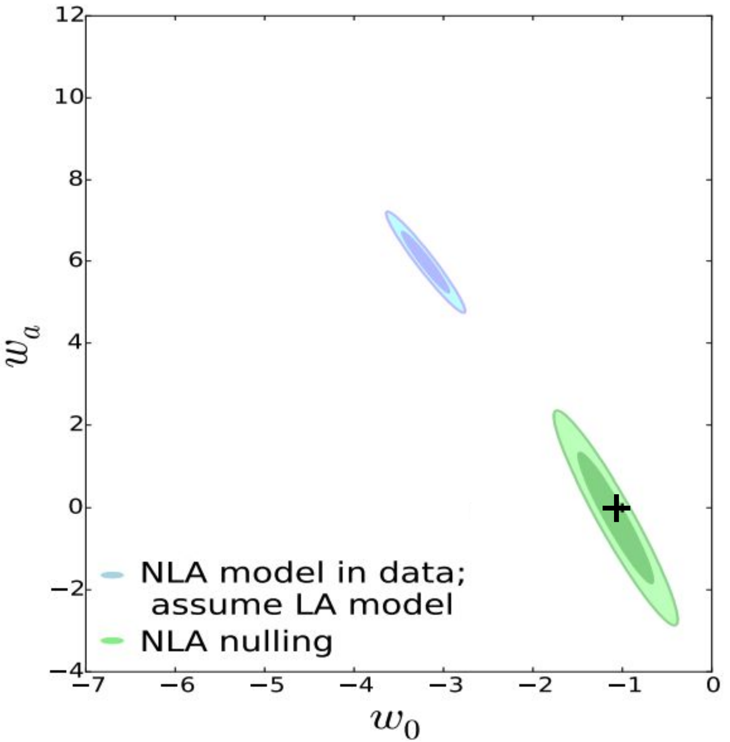
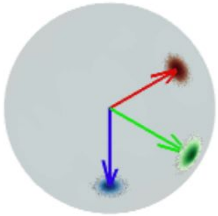


Image credit: Kirk et al, 2015

- Intrinsic Alignment (IA) is the tendency of galaxies to align with their neighboring galaxies and the underlying large scale structure
- This effect can masquerade as a weak gravitational lensing signal
- Need to develop good IA models
- Need to include realistically complex IA in the catalogs

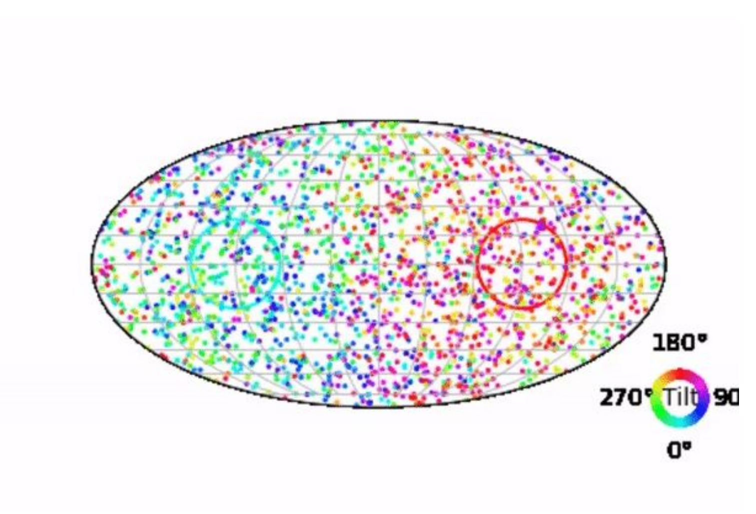
Diffusion:	SO(3)	Euclidean
		
• Noising Process		
• Noising kernel	Heat Kernel	Gaussian Distribution
• Score estimation	Lie derivatives	Gradients

• Score Matching

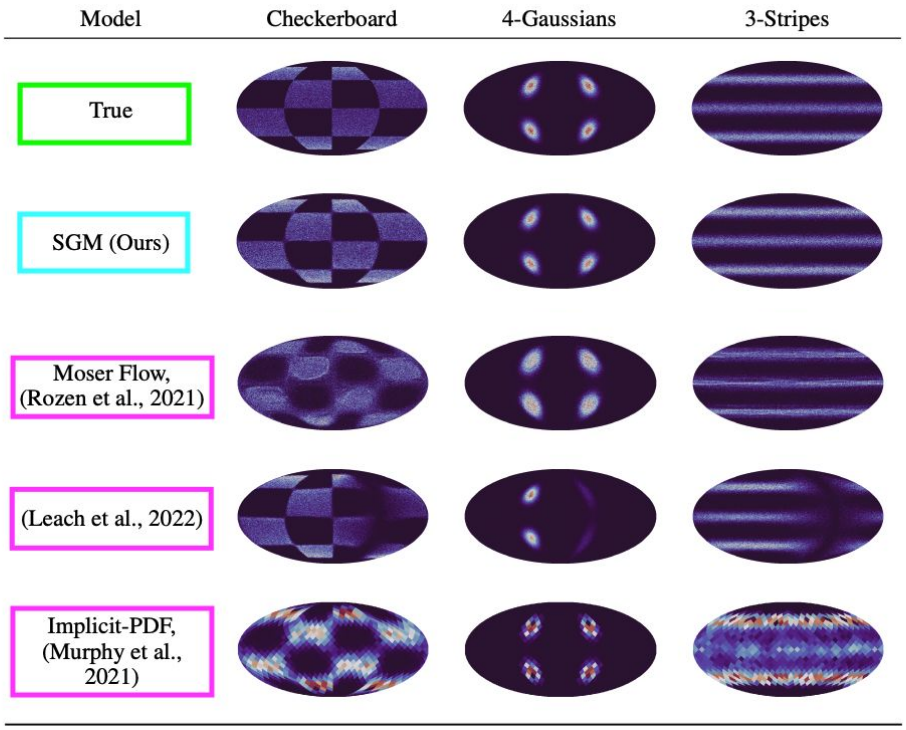
$$\mathcal{L}_{DSM} = \mathbb{E}_{p_{\text{data}}(\mathbf{x})}\mathbb{E}_{\epsilon\sim\mathcal{N}(0,\sigma_\epsilon^2)}\mathbb{E}_{p_{|\epsilon|}(\tilde{\mathbf{x}}|\mathbf{x})}\left[|\epsilon| \;\| \, s_\theta(\tilde{\mathbf{x}},\epsilon) - \nabla_X \log p_{|\epsilon|}(\tilde{\mathbf{x}}|\mathbf{x}) \,\|_2^2\right]$$

Credit: Gilitschenski et al, 2020

Successful Unconditional Density Estimation

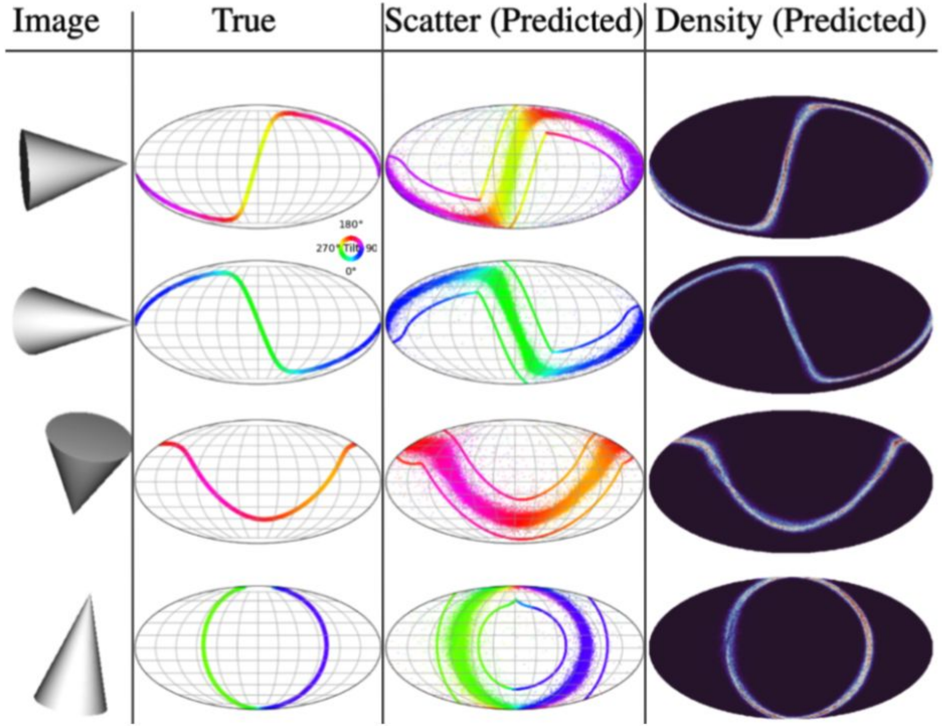


Toy example: mixture of two Gaussian blobs on SO(3)



Our SGM is the best-performing method of those considered for reproducing the true patterns in the top row

Diffusion on SO(3): Robust Conditional density

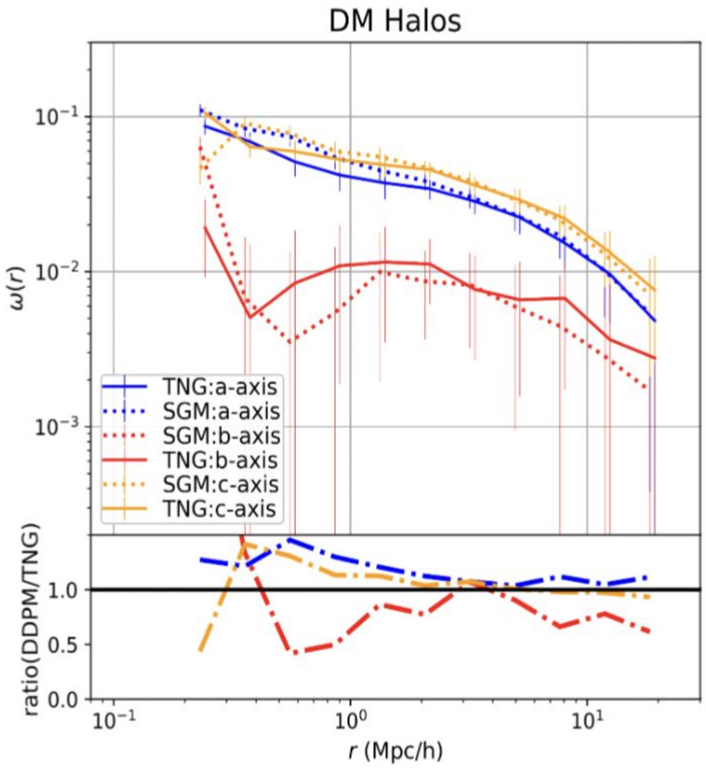


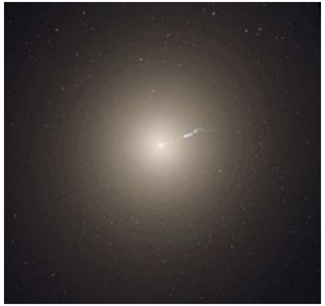
Computer vision/Robotics: pose estimation

Model can reproduce coherent galaxy alignments

The two-point correlation function, $\omega(r)$, which captures the correlations between galaxy axis directions and the positions of other galaxies.

Solid lines: measured values from the TNG simulation,
Dashed lines: generated values from the SGM, given **the tidal field**

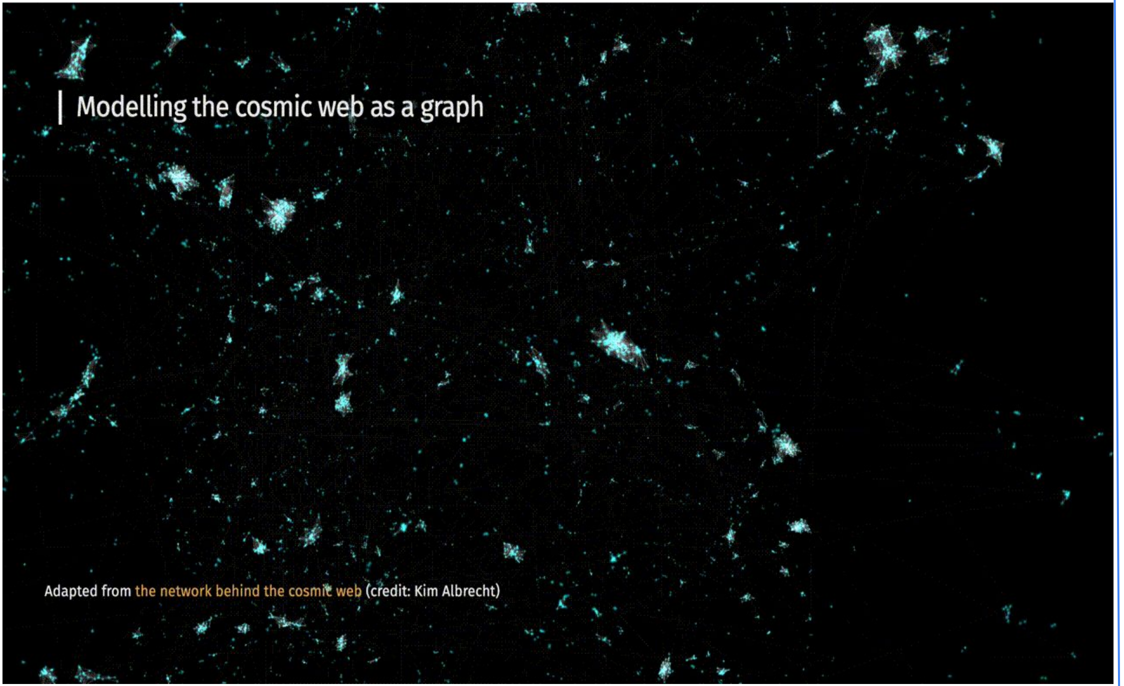




Galaxy orientation in 3D

arxiv:2212.05592

Current work: Adding (Equivariant) Graph layers



Cosmic Web as a set of Graphs

Our Contributions

A Deep Generative Model For Production of Synthetic Galaxies:

- We extend current SOTA diffusion onto the SO(3) manifold
- Showed robust applications in astrophysical context and computer vision
- Further work is needed to fully harness its power: **Equivariant Graphs,** for 1-halo regime
- Produce mock catalogs for Cosmological Surveys