Histogram Pooling Operators: An Interpretable Alternative for DeepSets

Miles Cranmer, Christina Kreisch
Alice Pisani, Francisco Villaescusa-Navarro, David N. Spergel, Shirley Ho
{mcranmer,ckreisch,apisani,fvillaescusa}@princeton.edu
{davidspergel,shirleyho}@flatironinstitute.org

ICLR 2021 Workshop
Deep Learning for Simulation (simDL)

PRINCETON UNIVERSITY

Overview

We demonstrate a differentiable histogram as a pooling operator in a DeepSet.

- For some set summarization problems, this results in much more interpretable latent features than standard approaches, due to functional resemblance with manual set summarization
- Allows us to extract explicit symbolic interpretations

We motivate and test this proposal with a large-scale summarization problem for cosmological simulations: predicting global properties of the universe via a set of observed structures distributed throughout space.

- The use of a DeepSet increases accuracy of traditional forecasting techniques from 20% to 13% for our dataset
- Histogram pooling achieves similar performance to sum- and mean-pool operations
- However, the histogram pool allows us to symbolically discover an optimal cosmological feature for cosmic voids, which is possible due to the strong connection with traditional pooling operators

Cosmic Voids Dataset

We use 2000 simulations from the Quijote simulation suite, each with ~5000 voids (Villaescusa-Navarro et al., 2020).

- Each void has features, such as radius, ellipticity, and depth

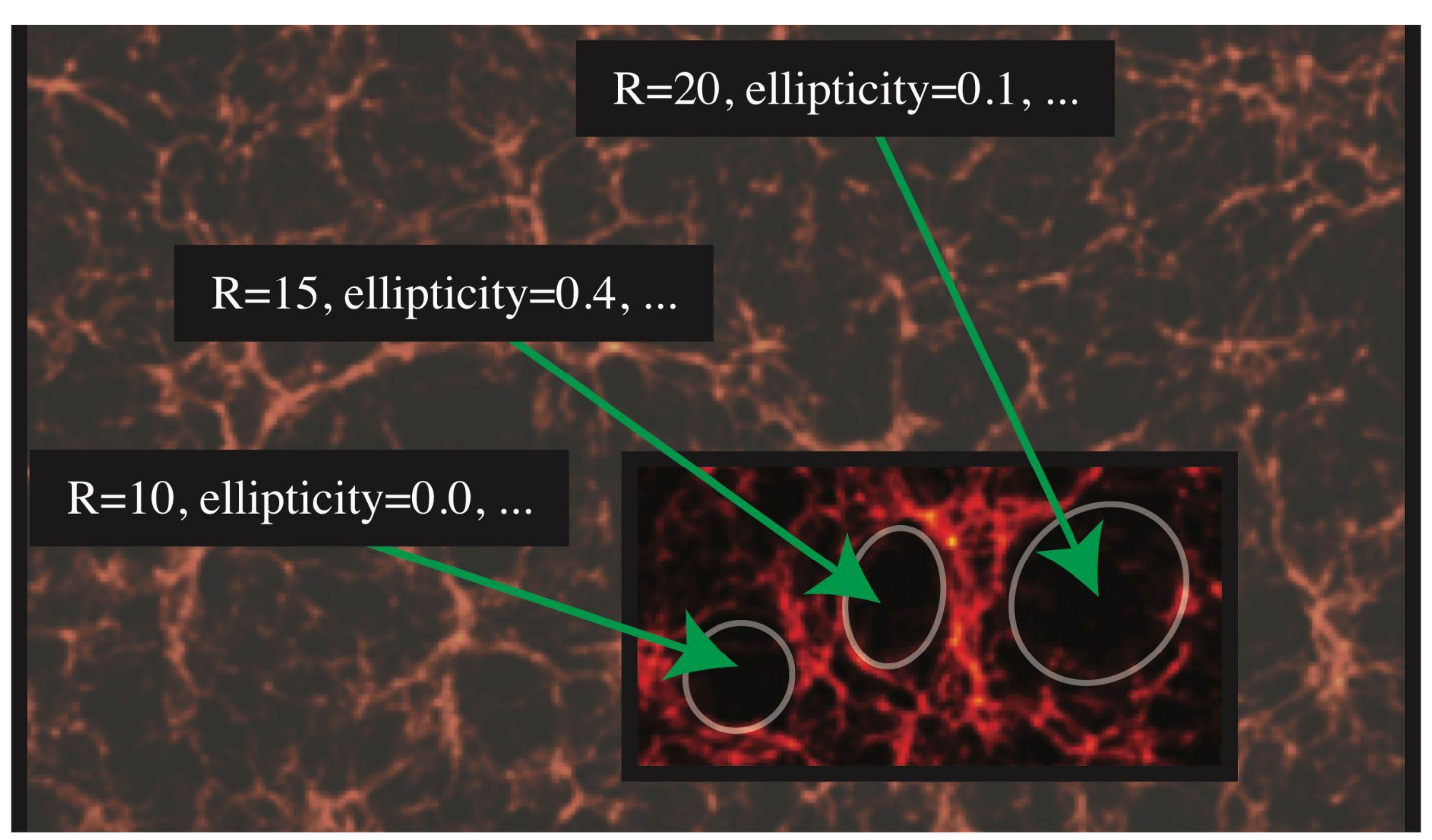

Figure 1: Field in the Quijote Simulations (Villaescusa-Navarro et al., 2020) with a few voids circled and labeled.

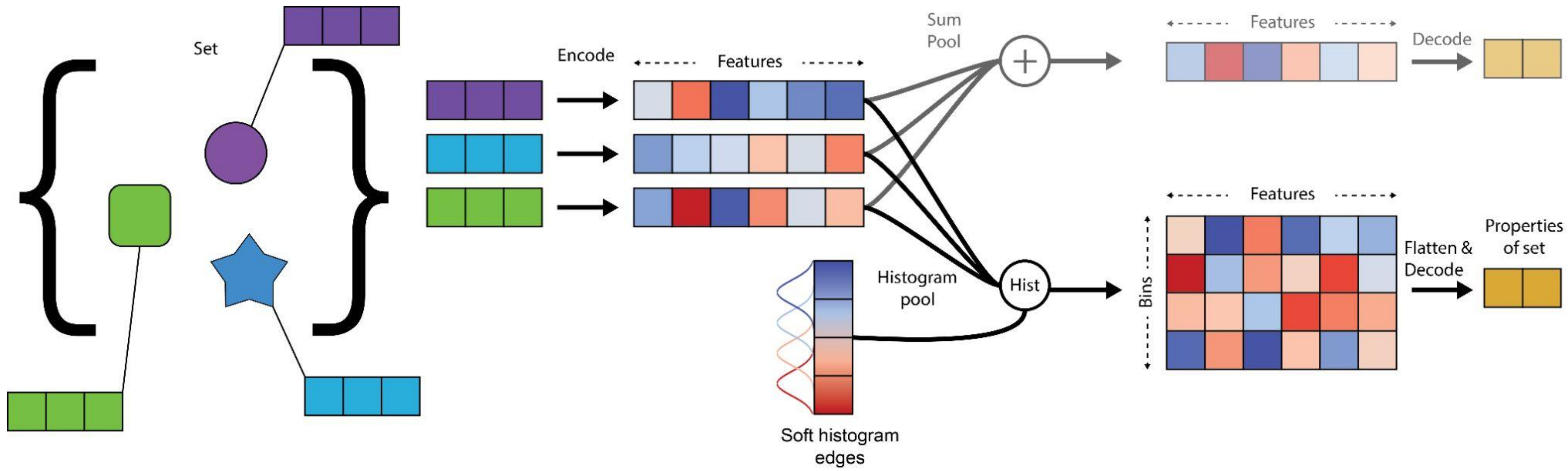

Figure 2: Schematic of a histogram pooling operator being used in a DeepSet, compared to a typical sum pool.

Model

A simple functional form for a DeepSet is:

$$\boldsymbol{y} = f(\,\rho(\{g(\boldsymbol{x}_i)\})\,)$$

for a set of vectors $\{x_i\}_{i=1:N}$, permutation-invariant pooling operator ρ, learned functions f and g, and output vector y.

Typical pooling operations include summation, averaging, or max. Our proposed histogram pooling operation takes the form:

$$\boldsymbol{y} = f(\text{flatten}(\boldsymbol{w})) \quad \text{where}$$

$$w_{jk} = \sum_i e^{-(a_k - z_{ij})^2/2\sigma^2}, \quad \text{for} \quad \boldsymbol{z}_i = g(\boldsymbol{x}_i)$$

where z_{ij} is the j-th latent feature of element i, a_k is a hyperparameter giving a pre-defined bin position for bin k, σ is a hyperparameter controlling the histogram's smoothness, w_{jk} is the histogram value for feature j and bin k, and $\boldsymbol{w}$ is a matrix with its j-th row and k-th column as w_{jk}. See Figure 2.

Results

Our best model yields the following interpretable equation for g:

$$z_{i1} = -\alpha R_i + \beta\delta_{ci} - \gamma R_i \epsilon_i + C,$$

$$\alpha = 0.17,\ \beta = 0.57,\ \gamma = 0.026,\ \text{and}\ C = 0.16$$

where R, □, and ϵ represent the void radius, depth, and ellipticity. Our model achieves ~13% error on predictions for Ω_m, improving over classic results of 20%, and allows us to have explicit analytic interpretations.

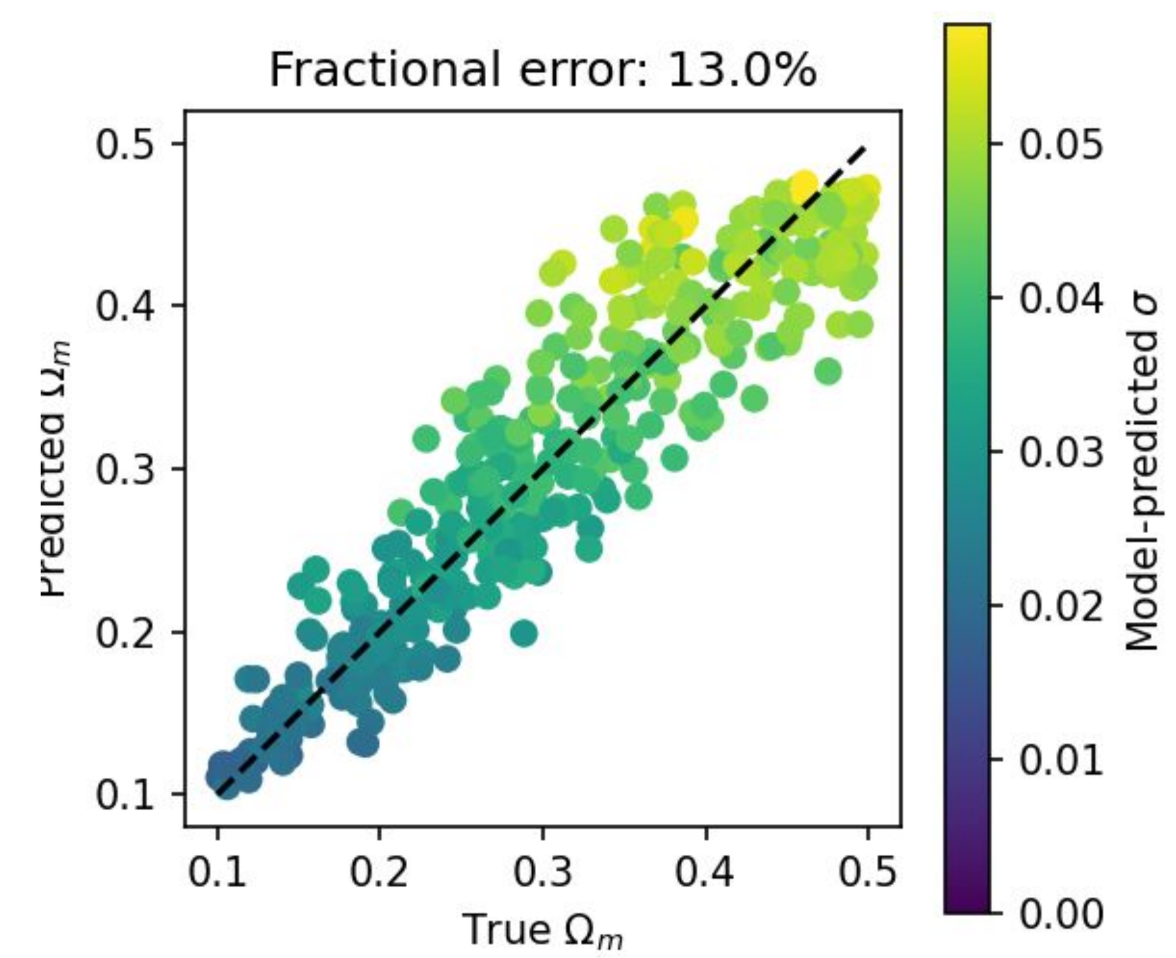

Figure 3: Value and error estimates for Ω_m using the best trained histogram model.

References
Villaescusa-Navarro F., et al., 2020, The Quijote Simulations, The Astrophysical Journal Supplement Series, 250, 2