Contribution ID: **13** Type: **Flashtalk with Poster**

Out-of-Distribution Multi-set Generation with Context Extrapolation for Amortized Simulation and Inverse Problems

Addressing the challenge of Out-of-Distribution (OOD) multi-set generation, we introduce YonedaVAE, a novel equivariant deep generative model inspired by Category Theory, motivating the Yoneda-Pooling mechanism. This approach presents a learnable Yoneda Embedding to encode the relationships between objects in a category, providing a dynamic and generalizable representation of complex relational data sets. YonedaVAE introduces a self-distilled multi-set generator, capable of zero-shot creating multi-sets with variable inter-category and intra-category cardinality, facilitated by our proposed Adaptive Top-q Sampling. We demonstrate that YonedaVAE can produce new point clouds with cardinalities well beyond the training data and achieve context extrapolation. Trained on low luminosity ultra-high-granularity data of Pixel Vertex Detector (PXD) detector at Belle II with $O(100)$ cardinality, YonedaVAE can generate high luminosity valid signatures with $O(10^5)$ cardinality and correct intra-event correlation without exposure to similar data during training. Being able to generalize to OOD samples, YonedaVAE stands as a valuable method for extrapolative multi-set generation tasks and inverse problems in scientific discovery, including de novo protein design, Drug Discovery, and simulating geometry-independent detector responses beyond experimental limits.

Primary author: HASHEMI, Hosein (ORIGINS Cluster)

Presenter: HASHEMI, Hosein (ORIGINS Cluster)