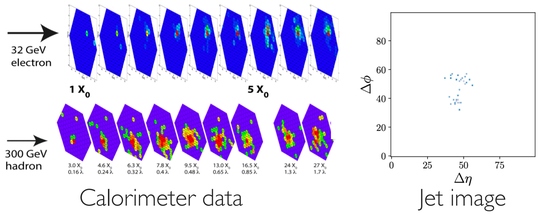
Graph Generative Adversarial Networks for Sparse Data Generation in High Energy Physics

Raghav Kansal[1], Javier Duarte[1], Breno Orzari[2], Thiago Tomei[2], Maurizio Pierini[3], Mary Touranakou[3, 4], Jean-Roch Vlimant[5], Dimitrios Gunopoulos[4]

[1]UC San Diego, [2]Universidade Estadual Paulista, [3]CERN, [4]National and Kapodistrian University of Athens, [5]California Institute of Technology

Particle Collision Simulations

- Simulations of collision events are important for data analysis in high energy physics
- Classical physics simulation programs are slow - O(min)/event
- Generative ML models may simulate at much higher speeds - O(ms)/event
- Current work on this involves linear or convolutional architectures
- These may not be well suited to the **sparsity** and **irregular geometry** of high energy data



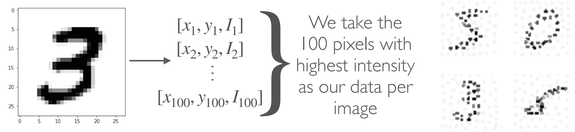
- Linear/CNNs need to represent jets and detector data as images

- Graphs naturally suit the sparsity of HEP data and adapt to any geometry
- So we aim for a **graph-based generative model**

Datasets

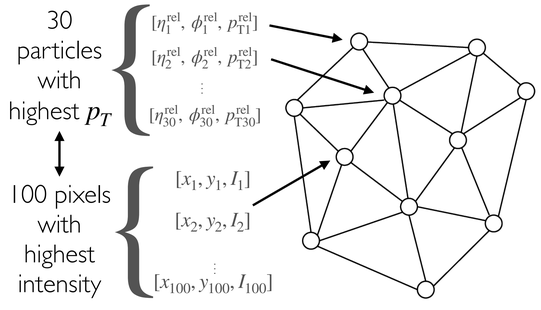
- We tested on two MNIST-derived datasets

Sparsified MNIST
(in analogy to jets or detector data)



MNIST Superpixels
(Monti et al. (2016))

- And a dataset of gluon jets - taking 30 highest p_T particles and their $[\eta^{rel}, \phi^{rel}, p_T^{rel}]$ features
- All three datasets are represented as graphs:



Evaluation

- We use a new Graph Frèchet Distance (GFD) metric (FID inspired) for MNIST datasets

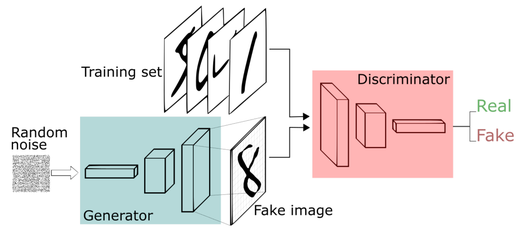
$$\mathrm{GFD} = ||\vec{\mu}_r - \vec{\mu}_g||^2 + \mathrm{Tr}(\vec{\Sigma}_r + \vec{\Sigma}_g - 2(\vec{\Sigma}_r\vec{\Sigma}_g)^{1/2})$$

μ's, Σ's are means and covariances of the activations of a pre-trained graph classifier on real and generated graphs

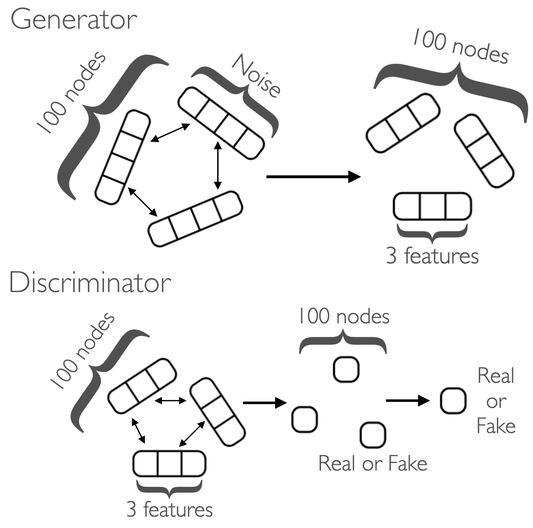
- For jets the 1-Wasserstein (W_1) distance between meaningful feature distributions

Graph GAN

- A GAN trains a discriminator and generator network iteratively and adversarially



- We use a **message passing neural network** architecture for both networks:

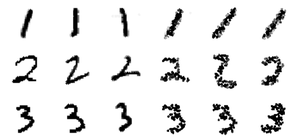


Summary

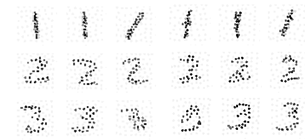
- We develop a novel Graph GAN
- Useful for sparse data and irregular geometry
- It's been successful on two MNIST-derived datasets and a jets dataset
- Future work will involve further applications to physics datasets, such as calorimeter data

Results

Sparse MNIST

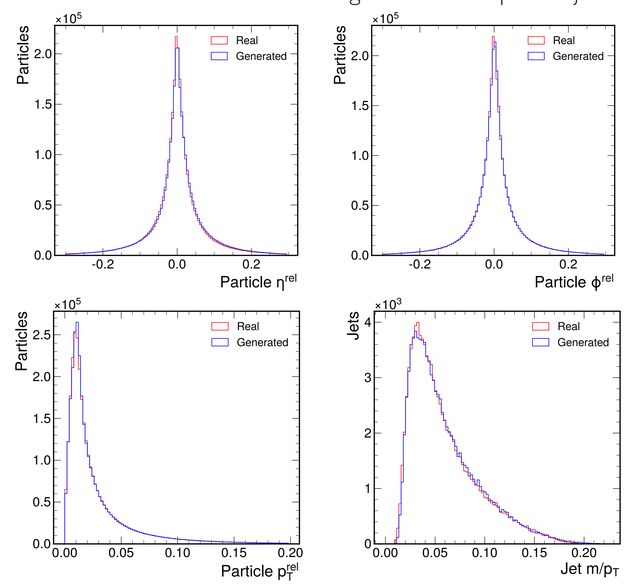


MNIST Superpixels



- We achieved an average GFD of 0.52 and 0.30 respectively

Feature distributions of real and generated samples of jets



W_1 scores between pairs of 100 jets

	η^{rel}	ϕ^{rel}	p_T^{rel}	Jet m/p_T
real vs real W_1 ($\times 10^{-3}$)	6 ± 2	6 ± 2	1.4 ± 0.5	6 ± 2
real vs gen W_1 ($\times 10^{-3}$)	5 ± 2	11 ± 4	2 ± 1	6 ± 2

- Real vs generated scores all within 1SD of real vs real scores

Details can be found in our paper (2012.00173)