Fast Simulation of Muon Background using Generative Networks - GANs

A. Marshall K. Petridis N. Serra

September 27, 2018

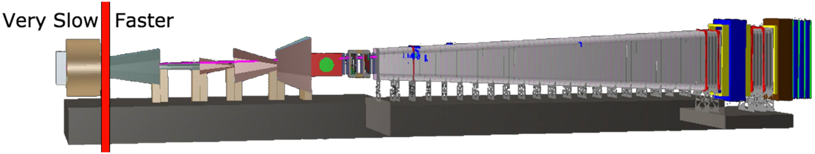
HH Wills Physics Laboratory
University of Bristol

This Talk

- Our goal
- What are Generative Adversarial Networks (GANs)?
- Network architecture and pre-processing
- Generated output
- Running generated muons in FairSHiP

Introduction to Problem

- Muon induced background studies would benefit from even larger simulation samples.
  - Simulated only $\sim 1s$ of full SHiP experiment.
- Larger sample particularly useful for muon-DIS background.
- Cannot currently generate anywhere near $\mathcal{O}(10^{20})$ POT.



Our Goal

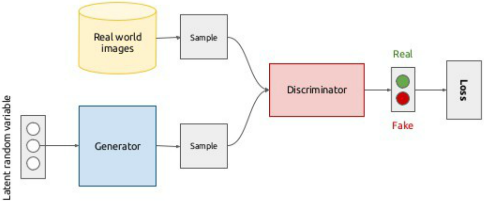
- SHiP muon background simulation is completed in two steps.
  1. Computationally expensive Pythia8 and Geant4 simulation of target and hadron absorber with charm cascade file.
  2. Muons that reach end of the hadron absorber in step 1 are re positioned at their point of production with correct kinematics in the full detector simulation.
- Our approach is similar. We attempt to model muons at the start of their tracks. The idea is to quickly generate new muons, muons not in the current simulated sample but muons that have parameters obeying these simulated distributions.
- Ultimate goal is to generate orders of magnitudes larger samples. Clearly this will come at some small expense of the fidelity of the generated muons

Generative Adversarial Networks (GANs)

- Modern machine learning technique, GANs are employed mostly in the generation of fabricated images that obey characteristics of a training set.
  - Examples in machine learning literature have high resolution generated images indistinguishable to training set.
- GANs currently very popular in ML, outside applications are becoming more common.
- So far just a few implementations in HEP. This will grow!

Generative Adversarial Networks (GANs)

- Basic idea is,
  - Two neural networks, a generator G and a discriminator D.
  - G takes in latent random noise, and outputs fabricated images.
  - D takes in images and predicts whether they came from a training sample or are fakes generated by G.
  - D begins to accurately understand the characteristics of the real images, which span a large dimensionality space.
  - G learns from the output of D to produce images that are ever-improving representations of the training sample in an iterative (step) procedure.

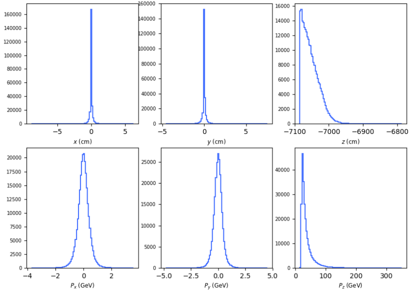


Our Approach

- We extract start positions and momenta values for muons that reach the end of the hadron absorber from Thomas' target simulation.
- We keep information about each muon's weight (due to enhancement of vector meson decays and γ conversion).
  - During training, these weights are used to negate this enhancement i.e. currently we model muons according to physical production fractions.

Pre-processing 1

Distributions of each parameter in the training data:



- Split full training set into μ^+ and μ^-.
- For each muon create an 'image' from the 6 parameters $[StartX, StartY, StartZ, P_x, P_y, P_z]$.
- Map 'pixels' to values between -1 and 1, saving min/max values from sample for later use in generation.

Pre-processing 2

Additional steps, firstly:

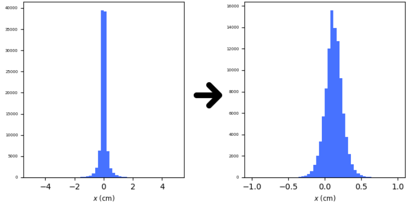
- Turns out GANs can struggle (takes longer for convergence) to model highly discontinuous features (see *StartX* and *StartY* distributions).
  - In the target simulation $StartX = 0$ and $StartY = 0$ about 1/3 of the time.
- To achieve better results we split the μ^+ and μ^- samples once again.
- For muons with $StartX = 0$ and $StartY = 0$ we create images with 4 features, [$StartZ$, P_x, P_y, P_z].

Category	Particle	Saved Properties					
1	μ^+	$x \neq 0$	$y \neq 0$	z	p_x	p_y	p_z
2	μ^+	$x = 0$	$y = 0$	z	p_x	p_y	p_z
3	μ^-	$x \neq 0$	$y \neq 0$	z	p_x	p_y	p_z
4	μ^-	$x = 0$	$y = 0$	z	p_x	p_y	p_z

Pre-processing 3

Secondly:

- For the 6 feature case, the *StartX* and *StartY* distributions are broadened in a controlled and reversible manner using the sqrt distance from mean value.
- This feeds the GAN a distribution it can more easily model.
- After generation, GAN output values are converted back to real values.



GAN Architecture

- Both G and D in the GAN are built from fully connected dense layers.
- I run tests on GPU cluster at Bristol to find optimal hyperparameters (number of layers, number of nodes in each layer, learning rate, choice of optimizer etc)
- Best found solutions have D with same structure for 4 and 6 pixel GAN. 2 hidden layers in an inverted pyramid structure. G has 2 hidden layers for both cases but significantly more nodes in the 6 pixel case.
- Training takes $\sim$ 60hrs for the 6 pixel case, and $\sim$ 20hrs for the 4 pixel case on single NVIDIA Tesla P100 GPUs.

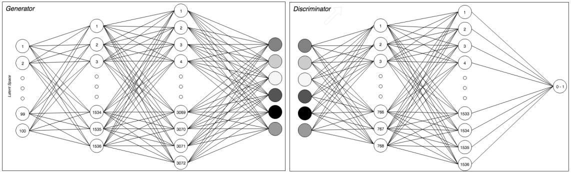
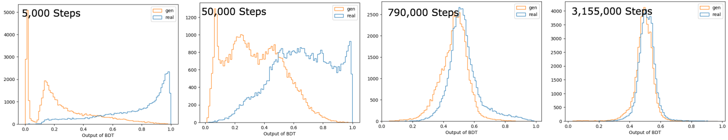


Figure of Merit

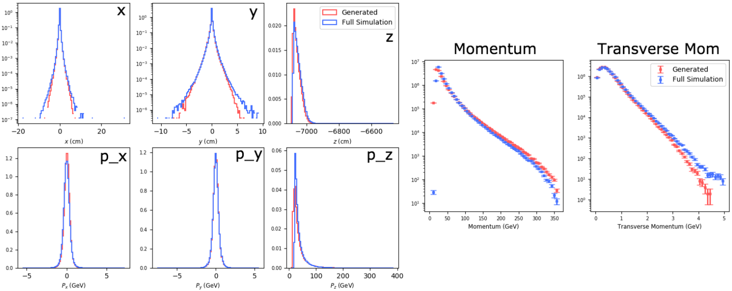
- To test the model quantitatively every 5000 steps we train a BDT to distinguish between a labeled training and a labeled generated sample.
- 50k real and 50k generated test images are then fed through the BDT and output plotted.
- This reduces the dimensionality of the images and encodes all complex correlations between parameters.



- Alignment here means good GAN performance.
- We track χ^2 values between these two distributions throughout training the GANs, training is stopped when a long plateau is reached.

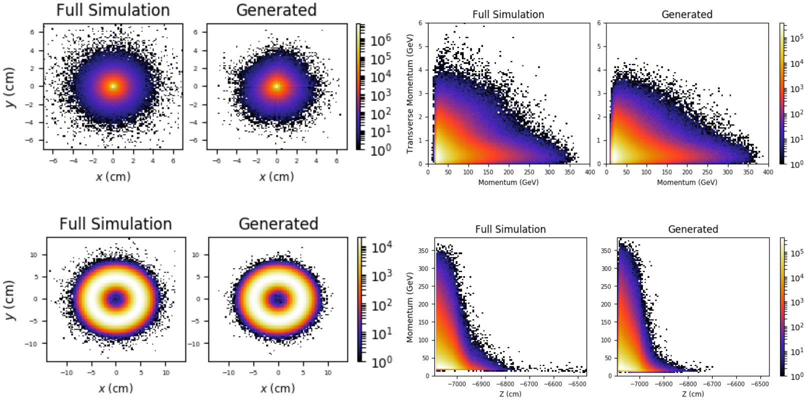
Generated Output Comparison

- We combine the output of the 4 GANs into one sample in physically correct ratios.
- Speed Up: On single GPU, GAN generates 1×10^6 muons in 90s. Speed up of $\mathcal{O}(10^6)$ relative to target simulation.



Generated Output Comparison 2D

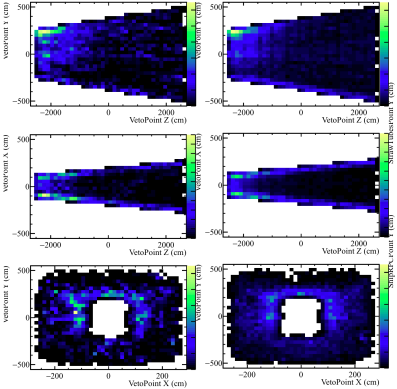
- Physically relevant 2D histograms show the GANs ability to accurately represent complex correlations between parameters.



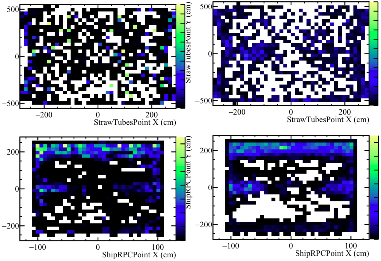
Running on FairSHiP

Example GAN here has $\sim 3.5 \times 10^{11}$ POT (run without beam smearing).
Currently feed with a custom generator.

Full Simulation GAN



Full Simulation GAN



Plots are correctly weighted, normalized and colour scales are matched.
PLOTS ARE OLD GAN MUONS, CURRENT GAN IS MORE ACCURATE.

Future

- Need a reconstructed track momentum plot from tracks in strawtubes for GAN vs full simulation muons. Will have this soon.
- Using data from muon test beam to train GAN?
- We are in process/planning to write a dedicated paper on this novel application of GANs.

Summary

- GANs extremely well model the phase space covered by the full simulation output (training sample).

- This method can provide a huge speedup $\mathcal{O}(10^6)$ and massive increase in statistics for SHiP muon induced backgrounds.