(FAT-GAN) architecture for simulation of electron-proton scattering events

Yasir Alanazi, N Sato, Tianbo Liu, W Melnitchouk, Michelle P Kuchera, Evan Pritchard, Michael Robertson, Ryan Strauss, Luisa Velasco, Yaohang Li

Department of Computer Science, Old Dominion University, Norfolk, Virginia 23529
Theory Center, Jefferson Lab, Newport News, Virginia 23606
Department of Physics, Davidson College, Davidson, North Carolina 28035
Department of Mathematics and Computer Science, Davidson College, Davidson, North Carolina 28035
Department of Physics, University of Dallas, Irving, Texas 75062
Generative Adversarial Network (GAN)

- Generative Adversarial Networks (GAN)
  - Introduced by Ian Goodfellow et al. in 2014
  - Deep neural network architectures comprised of two nets
    - A Generator
    - A Discriminator
  - Both nets are trying to optimize a different and opposing loss function in a zero-sum game

- Potential of GAN
  - Can be trained to mimic any distribution of data
  - Create worlds eerily similar to our own in any domain

“The most interesting idea in the last 10 years in machine learning” – Yann LeCun
General Architecture of GAN

Generator

Discriminator

True Data

Fake Data

Loss

Noise
The Power of GAN

- Can be trained to mimic any distribution of data

- Applications
  - Artificial Arts
  - Virtual Reality
  - New Characters
  - Artificial Music
GAN-based Event Generators

- Learning from real electron-proton scattering data
  - Capture rich underlying distributions over data
    - Difficult to model using explicit parameters
- Faithfully reproducing particle reaction events
  - No assumptions on femtometer-scale physics theory
- Overcome the limitations of MCEGs
Problems in Training a GAN

- Training a GAN is notoriously difficult
  - Perfect Discriminator
  - Mode Collapse
  - Non-convergence
  - Imbalance Generator and Discriminator Training
  - Model parameter oscillation
  - Destabilization
  - Vanishing gradient
Too Perfect Discriminator

- $P_{G_0}(x)$ vs $P_{data}(x)$:
  - $JSD(P_{G_0} \| P_{data}) = \log_2$

- $P_{G_{50}}(x)$ vs $P_{data}(x)$:
  - $JSD(P_{G_{50}} \| P_{data}) = \log_2$
  - Not really better ......

- $P_{G_{100}}(x)$ vs $P_{data}(x)$:
  - $JSD(P_{G_{100}} \| P_{data}) = 0$
Mode Collapse

- Generating the same sample

Generated Distribution

Data Distribution

- 8 -
Additional GAN Challenges for Physical Event Generators

- Precise Event Feature Distributions
  - Replicate the nature of particle reactions faithfully

- Obeying the fundamental Physics Laws
  - Energy Conservation
  - Momentum Conservation
GAN for Inclusive Events

- Only learns how to generate specific kinds of particles
  - Instead of the whole spectrum of particles in one event

- Single electron

- Electron + Prion
Direct Simulation GAN

- Directly learn from electron three-momentum vector (px, py, pz)
Direct Simulation GAN (cont.)

- Inter-correlation between physical quantities
Momenta Distributions of Electrons

![Graphs showing momenta distributions for p_x, p_y, and p_z.](image)
Feature-Augmented Transformed GAN (FAT-GAN)

- **FAT-GAN**
  - Features Transformation
    - Select generated features
      - Not necessarily meaningful physics properties
    - Easier to be generated by the generator
  - Features Augmentation
    - Expand feature space
    - Improve sensitivity of the discriminator
  - Maximum Mean Discrepancy (MMD)
    - Improve Distribution Match
  - Wasserstein Loss
    - Reduce the chance of mode collapse
    - Enhance GAN convergence
Features Transformation

\[ T(p_z) = \log\left( E_b - p_z \right) \]

- Conversion to eliminate sharp edges
- Guarantee no generation of non-physical electrons
Features Augmentation

- Augment the Feature Space to improve the Sensitivity of Discriminator
Wasserstein GAN

- **Discriminator**
  - No longer a direct critic of telling the fake samples apart from the real ones
  - Tries to bring $x_G$ distribution closer to $x_{true}$ distribution

- **Wasserstein Loss**

\[
D_{loss} = E(D(x_{true})) - E(D(x_G)) \\
G_{loss} = -E(D(x_G))
\]
Wasserstein Distance vs. JSD

\[ W(P_{G_0}, P_{data}) = d_0 \]
\[ W(P_{G_50}, P_{data}) = d_{50} \]
\[ W(P_{G_100}, P_{data}) = 0 \]

\[ JS(P_{G_0}, P_{data}) = \log 2 \]
\[ JS(P_{G_50}, P_{data}) = \log 2 \]
\[ JS(P_{G_100}, P_{data}) = 0 \]
Maximum Mean Discrepancy (MMD)

- **MMD**
  - A kernel-based two-sample test to compare two distributions
  - Determine if the two sets of samples are drawn from different distributions

\[
\text{MMD}^2(\mathbf{p}, \mathbf{\tilde{p}}) = \mathbb{E}_{\mathbf{p}_a, \mathbf{p}_a' \sim P_p}[k(\mathbf{p}_a, \mathbf{p}_a')] + \mathbb{E}_{\mathbf{p}_b, \mathbf{p}_b' \sim P_{\tilde{p}}}[k(\mathbf{p}_b, \mathbf{p}_b')] - 2\mathbb{E}_{\mathbf{p}_a \sim P_p, \mathbf{p}_b \sim P_{\tilde{p}}}[k(\mathbf{p}_a, \mathbf{p}_b)]
\]

- \(k(\mathbf{p}_a, \mathbf{p}_b)\) is a Gaussian kernel
FAT-GAN Architecture

- **Discriminator Loss**
  
  \[ L_D = \left( \mathbb{E}[D(\hat{p})] - \mathbb{E}[D(p)] \right) 
  + \lambda \mathbb{E}_{\hat{p} \sim P_{\hat{p}}} \left[ (\|\nabla_{\hat{p}} D(\hat{p})\|_2 - 1)^2 \right] \]

- **Generator Loss**
  
  \[ L_G = -\mathbb{E}[D(\hat{p})] + \eta \text{MMD}^2(p, \hat{p}) \]
Comparison between Representations in Cartesian Coordinates and Spherical Coordinates

- **Representation in Cartesian Coordinates [FAT-GAN (Cart)]**
  - \(px\)
  - \(py\)
  - \(pz\)

- **Representation in Spherical Coordinates [FAT-GAN (Spher)]**
  - \(\theta\)
  - \(\Phi\)
  - \(E\)
Distributions of Generated Physical Properties
Features Inter-correlations

$Q^2$
$GeV^2$

- **Pythia**
- **FAT – GAN (Cartesian)**
  $\chi^2 = 1.58$
- **FAT – GAN (Spherical)**
  $\chi^2 = 19.54$
- **DS – GAN**
  $\chi^2 = 82.41$
FAT-GAN for Two Particles (Electron + Prion)
Summary

- **GAN-based Event Generator**
  - Challenge: Complicated patterns in physics property distributions
  - FAT-GAN Features are the KEY
    - Not necessary using meaningful physics properties as generated features
    - Use features that are easy to generate by the generator
    - Augment feature space to make the discriminator sensitive

- **Success so far**
  - We can model the inclusive electron with high precision
  - We can model the inclusive electron and pion with high precision
FAT-GAN Paper


Codes: https://github.com/JeffersonLab/FAT-GAN
Acknowledgements

We thank Jianwei Qiu for helpful discussions. This work was supported by the LDRD project No. LDRD19-13 and No. LDRD20-18.