Toward a generative modeling analysis of CLAS exclusive $2\pi$ photoproduction
Tareq Alghamdi, et al.
Old Dominion University, Department of Computer Science, USA
talg001@odu.edu

Introduction:

Two-Pion Photoproduction in Hadron Spectroscopy:

• Crucial process in hadron spectroscopy addressing fundamental questions like the "missing baryons" problem.
• Widely used to determine the spectrum of unstable excited states.

Challenges with Multiparticle Final States:

• Introduction of a third particle complicates the reaction description.
• Requires additional variables for accurate representation, leading to a higher dimensional phase space.

Data Complexity and Extraction Challenges:

• Limited statistics and unpopulated bins in multidimensional space from Jefferson Lab CLAS g11 experiment.
• Large statistical fluctuations result in significant uncertainties in extracting the underlying reaction mechanisms.

Motivation

• Particle Collision Experiments:
  – Involve particles like electrons and protons.
  – Accelerated to nearly the speed of light.
  – Collided to produce vertex level events.

• Detector Machines:
  – Used to observe and measure collisions.
  – Result in a set of events known as detector level events.

• Difference Between Vertex Level and Detector Level Events:
  – Due to imperfections and limited resolution of the detector.
  – Conceptualized as a folding process.
  – Represents a well-defined problem in physics.

• Unfolding Process:
  – Can we transform detector level events back into vertex level events?
  – Termined as "unfolding".
  – Inherently ill-posed and challenging.

• Machine Learning Approach:
  – Exploration of using machine learning to reconstruct vertex level events.
  – Represents a novel and promising direction.
  – Serves as a primary motivation for this research.

Event Generators in Particle Physics

• Monte-Carlo Based Generators:
  – Traditional approach for simulating physics events.
  – Constructed using a combination of experimental data and theoretical inputs defined by physics rules.
  – Well-known tools in the literature: Pythia, Herwig, etc.

• Machine Learning Based Generators:
  – Employ a data-driven approach to learn directly from event samples.
  – Mostly based on advanced machine learning techniques like GANs, Variational Autoencoders, and Normalizing Flows.

Challenges

To faithfully replicate the nature of particle reactions, precise modeling of event features and correlations is essential. The following challenges need to be addressed:

• Calculating Derived Physics Variables:
  • Variables not included in the training dataset: $E_\text{REC}$, $t$, $t'$, $\cos \theta$.

• Transferring Training Results:
  – Transfer training results of invariant variables to other spaces:
    • From Center-of-Mass (CM) frame to Laboratory (Lab) frame.

• Calculating Momentum Resolution:
  – Calculate momentum resolution $\Delta p = \sqrt{p_x^2 + p_y^2 + p_z^2}$ (features not included in training).

Methodology

• Developed an ML-based detector simulation using a conditional GAN.
• Trained a conditional GAN generator to simulate the detector’s smearing effect.
• Generated synthetic REC detector-level events from input noise and PS-MC GEN events.
• Passed GEN PS-MC accepted events through the GATE chain to obtain REC pseudodata.
• Concatenated both synthetic REC and REC pseudodata with original GEN events.
• Fed concatenated data to the GAN discriminator to facilitate convergence.
• During training phase, DS-GAN learned how the detector impacts the original data through event-by-event comparison of GEN and REC pseudodata.
• After successful training, DS-GAN generator serves as the ML detector surrogate.
• Integrated the DS-GAN into the UNF-GAN architecture.

• Demonstrated the use of a Generative Adversarial Network (GAN) to realistically reproduce multibody physics reactions.

Results

DS-GAN

Figure 3: DS-GAN results (A) Training variables. (B) The derived variables (not used in the training).

Figure 4: (A) Evaluate the smearing by considering all momenta $p_x, p_y, p_z$, (B) 2D distributions from MC data and GAN result.

UNF-GAN

Figure 5: UNF-GAN results. (A) Training variables. (B) The derived variables (not used in the training).

Figure 4: (A) Exploring other physics-relevant distributions not used in the training as well for variables $y$ momentum components and $\theta$ in lab frame. (B) 1D histograms for fixed slice of the other variables.

Uncertainty Quantification

• Importance of assessing reliability of ML predictions in physics.
• Systematic uncertainty from training procedure quantified using bootstrap resampling.
• DS-GAN: 20 neural networks trained independently on different random samples.
• UNF-GAN: Adopted similar training with 20 independent networks, each using a different DS-GAN.
• Combined systematic uncertainties from DS- and UNF-GANs.

Conclusions

• Demonstrated the use of a Generative Adversarial Network (GAN) to realistically reproduce multibody physics reactions.
• Demonstrated that the GAN can reproduce training and derived kinematic variables and successfully unfold detector effects (smearing) across multiple dimensions.
• We demonstrate that the original correlations are preserved.
• The unknown quantification of the entire procedure was assessed by combining a bootstrap for the two NNs.