

ABSTRACT

This Masters Dissertation outlines the application of deep learning methods on raw data from the Transition Radiation Detector at CERN as well as on simulated data from the Monte Carlo Event Generator Geant4, in order to achieve the following goals:

- i. Classification Part I: Particle identification, distinguishing between electrons and pions

To this end, various feedforward neural networks, convolutional neural networks, as well as recurrent neural networks were built using Keras with a TensorFlow back-end, resulting in an ultimate pion efficiency (based only on features created from TRD raw data) of $\varepsilon_{\pi} = 7.86 \times 10^{-7}\%$ at electron efficiency $\varepsilon_e = 89.99\%$. By supplementing this data with kinematic parameters η, θ and ϕ , pion efficiency was reduced to $\varepsilon_{\pi} = 7.37 \times 10^{-8}$ at $\varepsilon_e = 90.01\%$, and when adding dE/dX measurements from the TPC to the set of features, an ultimate pion efficiency of $\varepsilon_{\pi} = 0$ at $\varepsilon_e = 99.99\%$ was achieved on an independent test dataset of 246 355 observations.

Raw data was extracted from the Worldwide LHC Computing grid using the ROOT data analysis framework, a C++ based platform maintained by physicists at CERN. R and Python were used interchangeably during various stages of data exploration, processing, analysis and model-building.

- ii. Classification Part II: Distinguishing real data from data generated by Geant4

This stage of the project focused on employing convolutional neural networks towards distinguishing real data from simulated data. Data was simulated using Geant4, a Monte Carlo toolkit which simulates the passage of particles through matter. ROOT was used to reconstruct the simulated data to deliver it in a similar format to that given by raw data after processing. A balanced accuracy of 91.5% (with Sensitivity = 0.8575 and Specificity = 0.9725) was achieved, using a 2D Convolutional Neural Network.

- iii. Deep Generative Modeling: Prototyping Variational Autoencoders and various Generative Adversarial Networks towards data generation

Various deep generative models were built to take as input raw TRD data and produce simulated observations which are likely under the training data distribution. While results were not as accurate as simulations from Geant4 (VAE simulations could be distinguished from real data to 100% accuracy), it is nonetheless worthwhile to see VAEs and GANs performed on this type of problem.

THE APPLICATION OF DEEP LEARNING TECHNIQUES TOWARDS PARTICLE IDENTIFICATION AND
HIGH ENERGY PHYSICS EVENT SIMULATIONS RELATING TO THE ALICE TRD AT CERN

CG Viljoen (VLJCHR004) – MSc Data Science Thesis – PHY5008W