Event Classification at SoLid using Semi-Supervised Generative Adversarial Networks

Imperial College London

Robert Jones & Edward Palmer | Supervisor: Antonin Vacheret | High Energy Physics Group

Sterile Neutrinos

Sterile neutrinos are potential dark candidates that exist beyond the standard model. They interact only through the gravitational force making them hard to detect [1]. Their discovery would provide a bridge to beyond standard model physics.

In 2018, the MiniBooNE detector observed a 4.5σ excess of electron neutrino events between 200-1250 MeV which could be the result of sterile neutrinos with energies of order electron-volts. However, evidence for their existence is conflicted [2,3].

SoLid

"The Search for Oscillations with a Lithium-6 Detector"

- SoLid is a 1.6-ton neutrino detector installed 6m from the BR2 reactor core. It consists of 12,800 5cm^3 PVT detector cubes arranged in planes facing the reactor [4].
- The detector measures the electron antineutrino flux close to the reactor core through inverse beta decays (**IBD**s). This allows measurement of "missing flux" due to oscillation to sterile neutrino states.

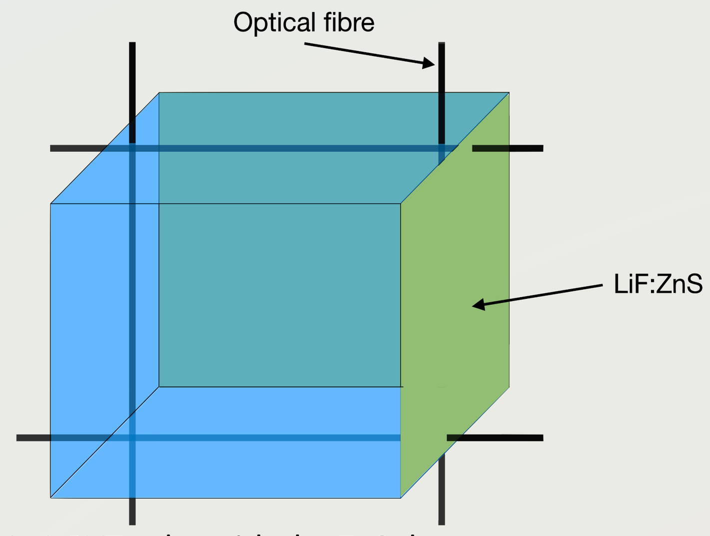

Fig. 1: A PVT cube with the ZnS detector screens fitted on two sides.

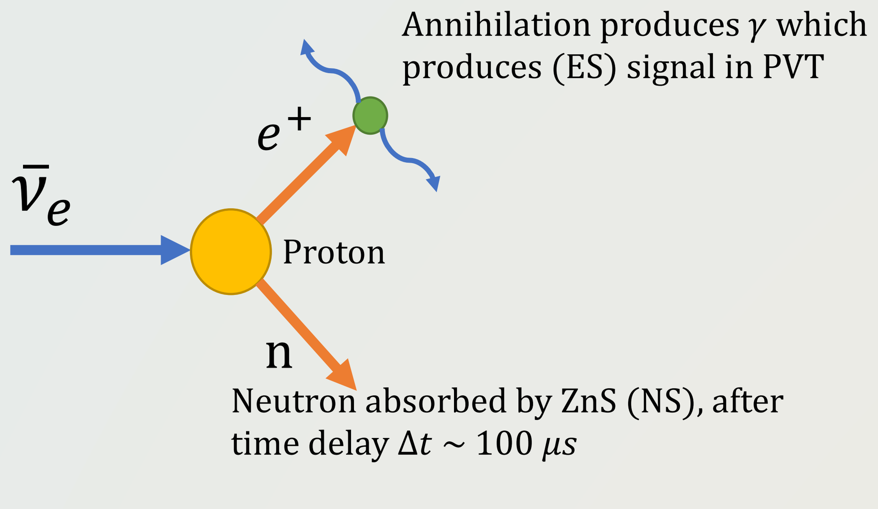

Fig. 2: A reactor antineutrino undergoes IBD producing a positron and neutron which are detected by the PVT cubes.

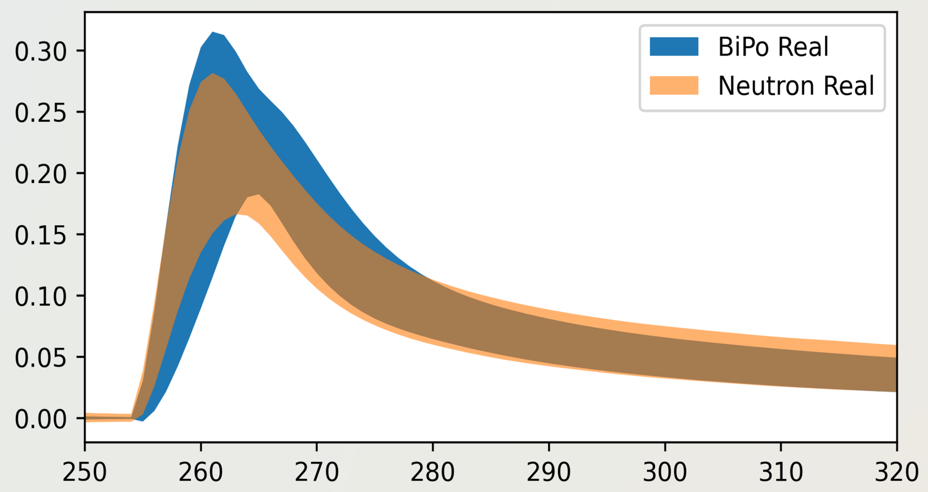

Fig. 3: A plot of the average real BiPo and neutron waveforms (±1 standard deviation).

- Events are filtered based on the time delay between **NS** (neutron scintillation) and **ES** (electron scintillation) signals. The remaining signals are then:
  1. **Accidental background:** Unconnected events which occur within the expected time interval.
  2. **Correlated background:** Decay of BiPo within the detector produces correlated signals with similar time delays to IBD events.
  3. **Signals:** IBD events caused by reactor antineutrinos (see fig. 2).
- Accidental background occurs at a constant rate so can be easily filtered out but BiPo events are harder to remove since they produce similar waveforms to IBD events.
- We focus on NS events to distinguish between BiPo and IBD signals.

Generative Adversarial Networks (GANs)

- GANs consist of two neural networks which are trained simultaneously with opposing loss functions. During training, the distribution of generated waveforms should approach that of the real distribution.
- During GAN training a chi-squared metric was used to quantify the difference between the normalized real and fake waveforms.
- We used schedulers to adjust the learning rate during training which allowed us to attain a lower chi-squared value and greater stability.

	Generator	Discriminator
Input	Batch of noise	1. Batch of fake waveforms 2. Batch of real waveforms
Output	Batch of fake waveforms	Batch of probabilities that waveforms are real
Goal	Fool discriminator into classifying fake waveforms as real	Distinguish real and fake waveforms

For each training iteration:
1. Pass real batch through discriminator and backpropagate loss through discriminator
2. Pass fake batch through discriminator and backpropagate loss through discriminator
3. Pass noise through the generator to generate a new fake batch
4. Classify fake batch using the discriminator and backpropagate loss through generator (maximise classification error)

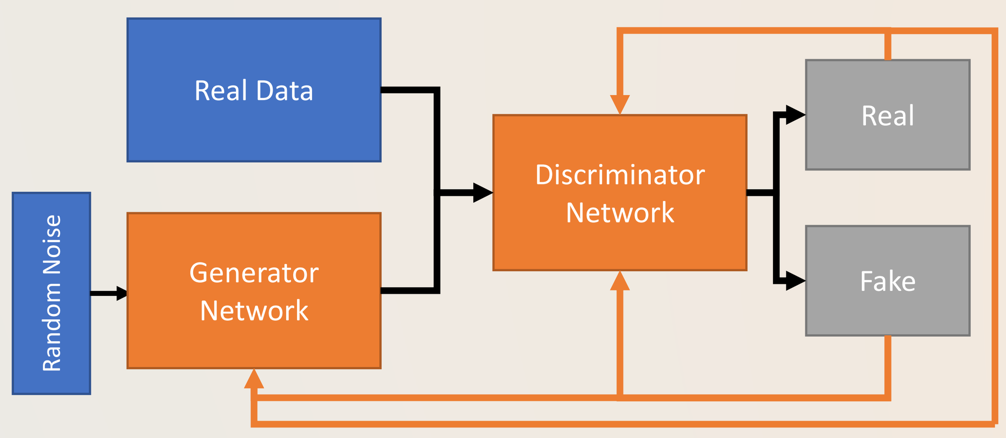

Fig. 4: Shows basic structure of a GAN where orange lines represent backpropagation steps.

Objectives

- Generate fake BiPo and neutron waveforms for use in simulations of the SoLid experiment.
- Improve the classification of BiPo events in the SoLid detector.

Semi-Supervised GANs (SGANs)

- The SGAN is developed from the GAN. It shares the same generator network and uses a core discriminator based on the GAN discriminator. The unsupervised and supervised discriminators share the core discriminator network but feature different final activation layers.
- Unlike GANs, most of the SGAN training dataset is unlabelled (90%).
- After training, the supervised discriminator can be used to classify real BiPo and neutron waveforms.

	Generator	Supervised Discriminator	Unsupervised Discriminator
Input	Batch of noise	Batch of labelled real BiPo/neutron waveforms	1. Batch of unlabelled real BiPo/neutron waveforms 2. Batch of unlabelled fake waveforms
Output	Batch of fake waveforms	Batch of probabilities that real waveforms are neutron	Batch of probabilities that waveforms are real
Goal	Fool unsupervised discriminator into classifying fake waveforms as real	Distinguish real BiPo from real neutron waveforms	Distinguish fake from real waveforms

For each training iteration:
1. Train supervised discriminator on real labelled batch
2. Train unsupervised discriminator on real unlabelled batch
3. Train unsupervised discriminator on fake batch
4. Pass fake batch through unsupervised discriminator and backpropagate loss through generator (maximise classification error)

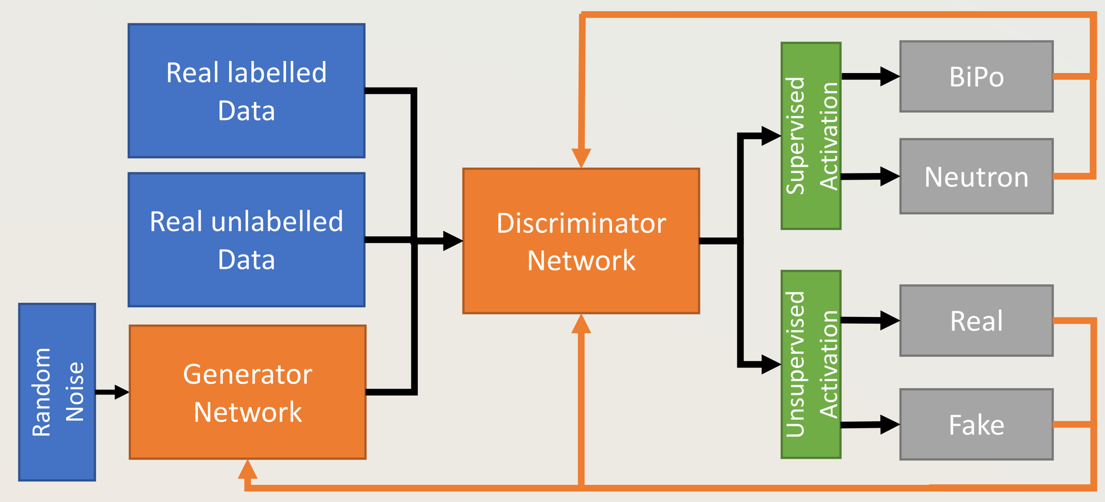

Fig. 5: Shows basic structure of an SGAN where orange lines represent backpropagation steps.

Preliminary Results

GAN

- After 50 epochs into training, the chi-squared metric is within 1 std of the mean metric calculated for the real batches.
- The average waveforms for both BiPo and neutrons were compared both visually and with the chi-squared metric for real and fake waveforms.
- Fig. 3 and fig. 6 show excellent agreement between the real and fake distributions.
- The generated waveforms can then be used to augment existing datasets and classifiers.

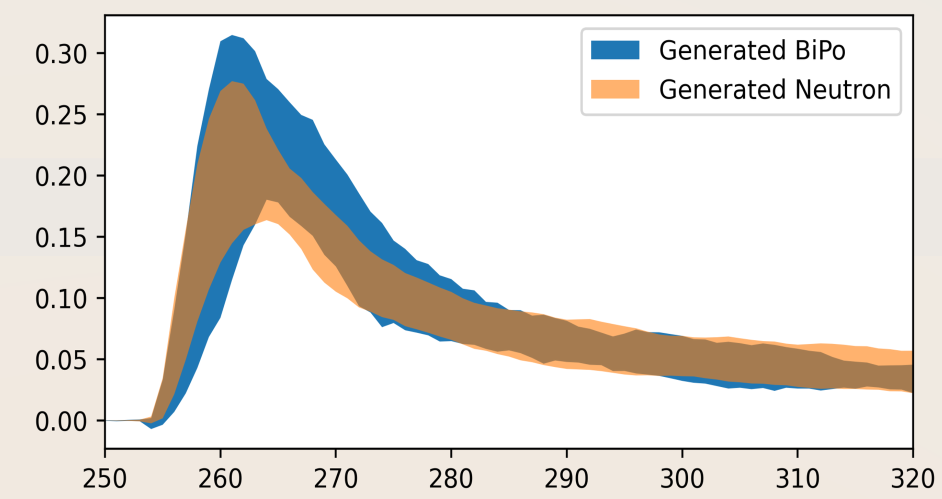

Fig. 6: The distribution of generated BiPo and neutron waveforms.

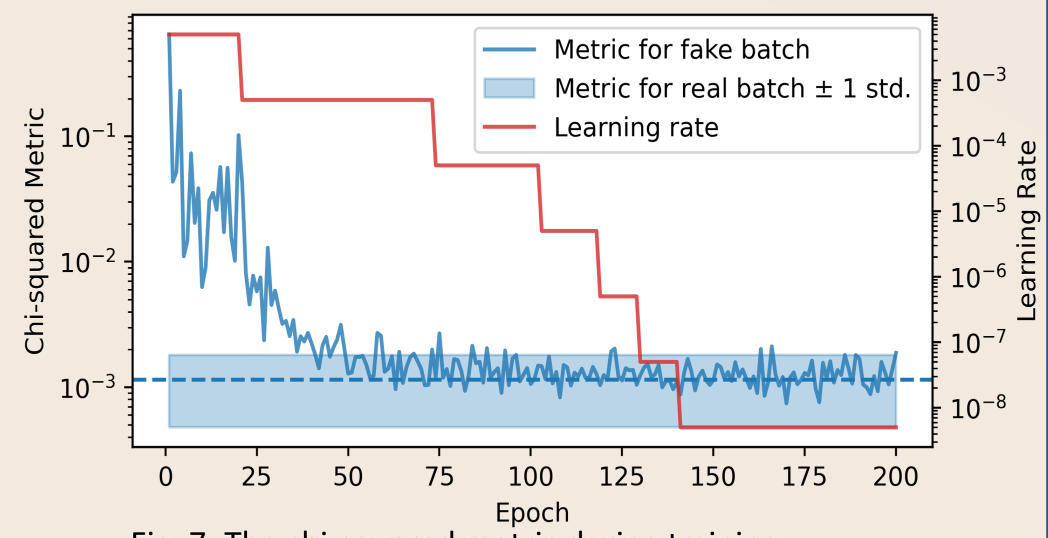

Fig. 7: The chi-squared metric during training.

SGAN

- The SGAN's accuracy on an unseen test dataset was computed every 50 iterations during training.
- The best results were achieved when training the SGAN on fake waveforms generated by our GAN in addition to real data.
- The SGAN discriminator achieved similar **AUC** scores to the previous best classifier, known as "BiPonator".
- AUC = area under an ROC curve. The ROC curve is a plot of true positive rate (TPR) against false positive rate (FPR) for a test dataset.
- There are multiple routes available to increase performance including:
  1. Adjusting the ratio of real to generated waveforms
  2. Adjusting the fraction of labelled waveforms
  3. Adjusting the hyperparameters

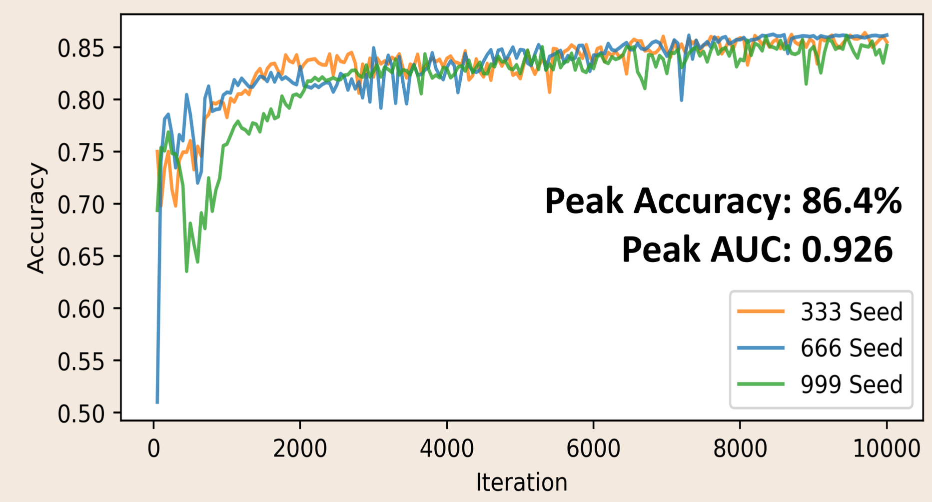

Fig. 8: Accuracy on an unseen test dataset during training.

- The SGAN's discriminator should be more robust than existing classifiers because it is also trained on generated data which enables it to cover a wider parameter space.
- The discriminator could be deployed to SoLid to better separate BiPo from real signals.
- This work will facilitate the ongoing search for sterile neutrinos.

Conclusion

- ☑ Successfully built a GAN to produce fake NS/BiPo waveforms that replicate the distribution of real waveforms.
- ☑ Successfully built an SGAN capable of correctly classifying NS/BiPo waveforms with an accuracy of >86%.

References

[1] (2021, August) All Things Neutrino—Fermi National Accelerator Laboratory. [Online]. Available: https://neutrinos.fnal.gov/types/ sterile-neutrinos/
[2] Aartsen, M., et al., 2016. Searches for Sterile Neutrinos with the IceCube Detector. Physical Review Letters, 117(7)
[3] Gariazzo, S., Giunti, C., Laveder, M. and Li, Y., 2017. Updated global 3+1 analysis of short-baseline neutrino oscillations. Journal of High Energy Physics, 2017(6).
[4] Y. Abreu, Y. Amhis, L. Arnold, G. Barber, W. Beaumont, S. Binet, I. Bolognino, M. Bongrand, J. Borg, D. Boursette, and et al., "SoLid: A Short Baseline Reactor Neutrino Experiment," Journal of Instrumentation, vol. 16, no. 02, pp. P02 025–P02 025, Feb 2021. [Online]. Available: http://dx.doi.org/10.1088/1748-0221/16/ 02/P02025