Punzi-loss

A differentiable metric approximation for sensitivity optimisation in the search for new particles

Huw Haigh[a], Paul Feichtinger[a], Gianluca Inguglia[a], James Kahn[b]
a Austrian Academy of Sciences Institute for High Energy Physics,
b Helmholtz AI, Karlsruhe Institute of Technology

HEPHY INSTITUT FÜR HOCHENERGIEPHYSIK

erc European Research Council Established by the European Commission

Introduction

The search for new particles requires robust classification methods, with the ability to be optimised for unknown cross sections and particle masses. We present a new loss function, *Punzi-loss*, based on the so-called Punzi figure of merit (FOM) [1]. We refer to a neural network trained with the Punzi-loss function as a *Punzi-net*, and investigate its application to the search for invisible decays of the hypothetical Z' boson produced in the process $e^+e^- \rightarrow \mu^+\mu^- Z'$ at the Belle II experiment [2].

Punzi-Loss Implementation

A figure-of-merit can be defined to describe the statistical significance provided when certain selection criteria are applied to data. A standard FOM used in particle physics is,

$$FOM = \frac{S(t)}{\sqrt{S(t)+B(t)}}. \quad (1)$$

Where $S(t)$ and $B(t)$ are the numbers of signal and background events surviving the selection (t). As the cross-section of a new process is unknown, standard particle physics FOMs are unsuitable. An alternative was proposed in [1], often referred to as the Punzi FOM. Using this, one can seek to maximise the inverse of the minimum detectable cross-section $\sigma_{\min}$,

$$\sigma_{\min}(t) = \frac{\frac{b^2}{2} + a\sqrt{B(t)} + \frac{b}{2}\sqrt{b^2 + 4a\sqrt{B(t)} + 4B(t)}}{\epsilon(t)\cdot L}, \quad (2)$$

We can build a differentiable function by replacing the fixed cut on the output with a sum over all events, weighted with the respective value of the output. If events classified as signal cluster around an output of 1 and events classified as background at 0, this quantity will closely approximate the original function. In Eq. 2 this weighting can be captured by performing the replacements

$$\epsilon(t) \rightarrow \epsilon(\boldsymbol{w},\boldsymbol{b}) = \sum_{\boldsymbol{x}} \frac{y_i \cdot \hat{y}_i(\boldsymbol{w},\boldsymbol{b}) \cdot s_{\rm sig}}{N_{\rm gen}} \quad \text{and} \quad (3)$$

$$B(t) \rightarrow B(\boldsymbol{w},\boldsymbol{b}) = \sum_{\boldsymbol{x}} (1-y_i)\cdot \hat{y}_i(\boldsymbol{w},\boldsymbol{b}) \cdot s^i_{\rm bkg}, \quad (4)$$

where the sum is over all training inputs $\boldsymbol{x}$ and the index i denotes the $i^{\rm th}$ training event. The collection of weights and biases that constitute the free parameters of the network are denoted as $\boldsymbol{w}$ and $\boldsymbol{b}$. Finally by summing the minimum detectable cross-section for each of the mass hypotheses, we yield the *Punzi-loss*;

$$C_{\rm Punzi} = \frac{1}{N_{Z'}} \sum_{m_{Z'}} \sigma_{\min}(\boldsymbol{w},\boldsymbol{b}), \quad (5)$$

variable	description
$p^*_{t,\rm thrust}(\mu)$	The transverse momentum component of the muons with respect to the thrust axis in the CMS.
$p^*_{t,\mu_{\min}}(\mu_{\max})$	The transverse momentum component of the higher energetic muon with respect to the lower energetic muon in the CMS.
$p^*_{l,\mu_{\min}}(\mu_{\max})$	The longitudinal momentum component of the higher energetic muon with respect to the lower energetic muon in the CMS.
$p^*_t(\mu^+\mu^-)$	The transverse momentum of the dimuon system in the CMS.

Table: The most important features found after training BDTs with many observables. These features are used for training the ANN.

Experiment

We trained a NN with 4 inputs, 2 hidden layers of 8 and 4 nodes respectively, and a single output node. The hidden layers utilise hyperbolic tangent activation functions while the output uses sigmoid. Training data comprised 1ab^{-1} of the main background process and 90 simulated Z' signals with masses in the range 0.1 - 9GeV, each comprising 20,000 events. The training of the Punzi-net was conducted in two steps;

1. The NN was first trained with the *binary cross entropy* (BCE) loss function, so as to introduce some initial separation between the signal and background distributions. It was found that without this the Punzi-loss could not converge.
2. The NN was then trained using the Punzi-loss, maximising the average Punzi FOM across the range of mass hypotheses.

Every second Z' signal is left out of training in order to be used for validation and a check for generalisation.

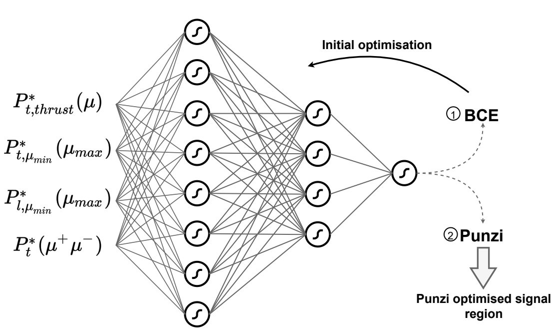

Figure: An outline of the network architecture.

Results

1. Improvement over BCE trained NN maximum achievable FOM in lower mass hypotheses.
2. Comparable result to BCE trained NN maximum achievable FOM in higher mass hypotheses.
3. Allows use of single cut to output of NN for all mass hypotheses, matching or outperforming the maximum FOM achievable by interpolated cuts to a BCE trained network output. This simplifies analysis workflow and studies of systematic uncertainties related to the use of a NN.
4. Good interpolation to all hypotheses not used in training.

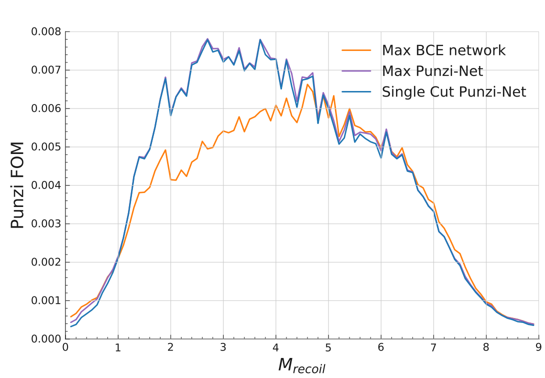

Figure: The Punzi FOM achieved by a single cut to the Punzi-net and maximum Punzi FOM achievable with the optimal varying cut to both the BCE trained network and Punzi-net for each mass hypothesis.

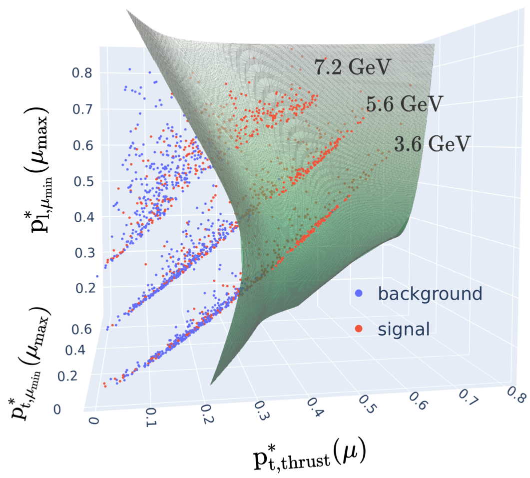

Figure: A 3D scatter plot showing the input space of the NN with $p^*_t(\mu^+\mu^-)$ fixed around 2.2 GeV/c. The separation boundary defined by the final selection (green sheet) separates the planes corresponding to different recoil masses in a way that optimises the selection for all signal hypotheses.

Conclusion and Outlook

1. We have demonstrated the *Punzi-loss* function, based on the Punzi FOM, and investigated its application in optimising the search for invisible decays of the hypothetical Z' boson.
2. The *Punzi-loss* function can bring improvements to the achievable FOM when compared to a NN trained with the more traditional BCE loss function.
3. In addition, the *Punzi-net* function allows for simplification of subsequent analysis since a common selection for all signal hypotheses can be applied to the classifier output.
4. These results represent a step towards a fully differentiable analysis framework in which optimisation of signal selection can account for systematic effects.

References

G. Punzi, Statistical problems in particle physics, astrophysics and cosmology. Proceedings, Conference, PHYSTAT 2003, Stanford, USA, September 8-11, eConf C030908, ArXiv: physics/0308063, 2003

Abe, T., Belle II Technical Design Report, Belle II Collaboration, KEK-REPORT-2010-1, arXiv:1011.0352, 2010

Acknowledgements

The authors would like to thank the Belle II Collaboration and Belle II software group for useful discussion.
P. Feichtinger, H. Haigh and G. Inguglia would like to acknowledge funding received under the Horizon 2020 framework of the European Research Council, ERC StG Nr. 947006 *InterLeptons*, and under the FWF standalone framework with grant Nr. P31361 *Searches for dark matter and dark forces at Belle II*.
James Kahn's work is supported by the Helmholtz Association Initiative and Networking Fund under the Helmholtz AI platform grant.