Neural Networks for the Gamma/ Hadron Separation of the Cherenkov Telescope Array

E. Lyard[1], R. Walter[1], V. Sliusar[1], N. Produit[1]
for the CTA Consortium[2]

[1]Department of Astronomy, University of Geneva, Switzerland,
[2]See www.cta-observatory.org

ABSTRACT

The Cherenkov Telescope Array (CTA) will be the largest ground-based gamma-ray observatory. CTA will detect the signature of gamma rays and cosmic rays hadrons and electrons interacting with the Earth's atmosphere. Making the best possible use of this facility requires to be able to separate events generated by gamma rays from the particle-induced background. Deep neural networks produced encouraging results, but so far there has been no evaluation of their performance for gamma/hadron separation with respect to well established approaches. In this paper we compare convolutional neural networks and a standard analysis technique, namely boosted decision trees. We compare the performance of the two techniques as applied to simulated observation data. We then looked at the Receiver Operating Characteristics (ROC) curves produced by the two approaches and discuss the similarities and differences between both.

Supported by: UNIVERSITÉ DE GENÈVE FACULTÉ DES SCIENCES — cadmos center for advanced modeling science — CSCS

INTRODUCTION

We evaluated the performance of deep convolutional neural networks (CNNs) compared to Boosted Decision Trees (BDTs) when applied to the signal extraction of CTA. For this comparison, we took the output of an EventDisplay analysis of simulated events from a Monte-Carlo production of CTA. We then compared the Signal/Background performance obtained from both approaches.
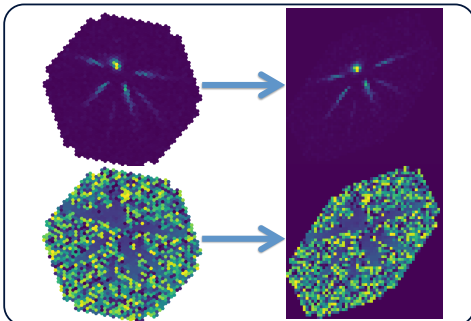

Fig. 1: Raw event data had its charge (top) and time-of-maximum (bottom) extracted. Charge was normalized while time-of-maximum remained untouched. Hexagonal images (left) were transformed into square images (right) so that we can work with standard CNN packages. This introduced a geometrical bias yet to be addressed.

METHOD

We used the exact same input for both the BDT and CNN approach. The standard BDT output is available to all CTA members by the CTA Analysis and Simulation working group while we performed the CNN analysis ourselves. Both approaches are compared using the ROC curve obtained from both analysis. Our CNNs operated on raw data, that we transformed as shown in Figure 1.

NETWORKS ARCHITECTURE

After some trial and error, we worked with 3 different architectures. The first had 18k parameters, the second had only twice the number of filters for each kernel, for a total of 290k parameters. Finally, a third version had the same number of filters as the second one, plus extra convolutions before the softmax layer (595k parameters). Architectures 2 and 3 can be seen in Figure 2. No dropout layer was included in the models.
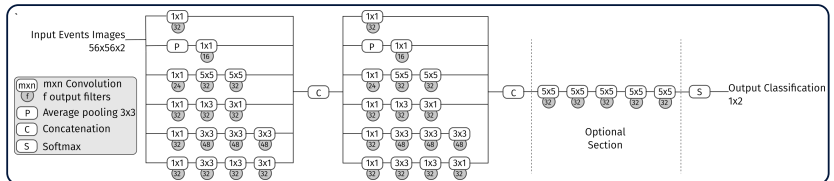

Fig. 2: CNN architectures used for this study.

RESULTS

Input data was classified per energy band, as is usually done for the existing analysis packages. ROC curves for two energy bands are presented in Figure 3. Overall CNNs performed in a very similar way to BDTs. CNNs outperformed BDTs at high energies, while the inverse was true at low energies. The actual ratio of signal/background events is on the order of 1/10000. It is thus very important that the ROC curve be as steep as possible to limit the contamination of the signal by background events. In some cases, even if the third network (red curve) tops out highest, it is not as steep as the others. Consequently, the best approach will depend on where one applies the cuts to separate signal from background. Figuring out the best possible cut is always a trade-off between significance and sensitivity, hence different cuts might be better suited for specific science goals.
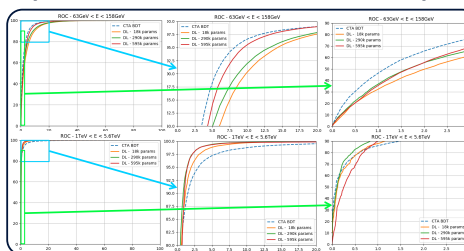

Fig. 3: ROC for energy bands 63-158 GeV (top) and 1-5.6 TeV (bottom). Left: zoomed out view, middle: zoom in the central region, right: zoom in the beginning region

CONCLUSION

These results are quite encouraging, as no a-priori knowledge was given to the CNNs. We demonstrated that CNNs can perform close to or already better than BDTs. This suggests that research in CNNs and other novel machine learning techniques should be actively pursued to help achieve the best science output for the upcoming CTA observatory.