

The use of Convolutional Neural Networks for signal-background classification in Particle Physics experiments

Venkitesh Ayyar^{1,*}, Wahid Bhimji¹, Lisa Gerhardt¹, Sally Robertson², and Zahra Ronaghi³

¹Lawrence Berkeley National Laboratory, 1 Cyclotron Rd, Berkeley, California 94720, USA.

²University of California, 101 Sproul Hall, Berkeley, California 94720, USA.

³NVIDIA, 2788 San Tomas Expressway, Santa Clara, California 95051, USA.

Abstract. The success of Convolutional Neural Networks (CNNs) in image classification has prompted efforts to study their use for classifying image data obtained in Particle Physics experiments. Here, we discuss our efforts to apply CNNs to 2D and 3D image data from particle physics experiments to classify signal from background.

In this work we present an extensive convolutional neural architecture search, achieving high accuracy for signal/background discrimination for a HEP classification use-case based on simulated data from the Ice Cube neutrino observatory and an ATLAS-like detector. We demonstrate among other things that we can achieve the same accuracy as complex ResNet architectures with CNNs with less parameters, and present comparisons of computational requirements, training and inference times.

1 Introduction

Particle physics experiments have been incredibly successful in improving our understanding of nature. An important aim of many of these experiments is to search for elusive particles with interesting properties. Evidence for these particles comes from rare events recorded in the detectors of these experiments, termed *signal*. These are accompanied by large number of other events from known and well-tested particle interactions, called *background*. To extract the interesting physics, it is essential to filter out the background from the signal. This process of distinguishing between the two is known as *Signal-background* classification and lies at the heart of experimental particle physics. Currently this is achieved by applying selections on derived high-level physics variables.

Given the success of deep learning methods in classifying real-world images, deep learning methods have been applied to the problem of signal-background classification ([1], [2] and references therein). In this work, we describe the use of Convolutional Neural Networks (CNNs) to classify signal from background for two use cases: a simulation dataset for the IceCube experiment and one for the ATLAS experiment.

2 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are a class of neural networks specialized for *image classification*. They are designed to capture features of images at different scales.

*e-mail: vpa@lbl.gov

Fig. 1 shows the basic structure of a CNN. It typically starts with a *Convolutional* layer: scanning through the input image in small blocks, it performs convolution operations to create feature maps. The dimensions of the resulting layers are large and they are reduced by the subsequent *Subsampling* layers. Stacking many such blocks sequentially can enable the networks to learn a variety of features at different scales. Eventually the neurons are combined together using one or more *Fully connected* layers. The output of final layer predicts the image class. CNNs have been very successful in classifying 2D images [3, 4].

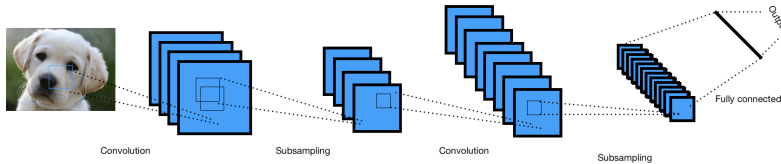


Figure 1: The general structure of a CNN with *convolutional*, *subsampling* and *fully-connected* layers stacked together.

3 Dataset 1: IceCube

3.1 The IceCube experiment

IceCube is a neutrino observatory located at the South Pole, looking for high energy ($>100\text{GeV}$) astrophysical neutrinos [5, 6]. Interactions of these high energy neutrinos with nuclei produce secondary charged particles. These emit Cherenkov light and are detected using an array of Digital Optical Modules (DOMs) placed below the ice. The experimental setup is depicted in Fig. 2a.

While different particles can contribute to the Cherenkov light seen by the DOMs, this dataset only considers the contribution of *muons*. The Cherenkov radiation from muons produced by astrophysical neutrinos forms the *signal* for this dataset. The *background* consists of the light contribution from other atmospheric muons. This dataset used the high energy down-going region of IceCube detection, a region not used in most analysis due to the high background component. The physical way to distinguish the signal and background in this region is using the stochasticity of energy deposition. Signal muons obtained from reactions of astrophysical neutrinos are single muons and hence lose energy stochastically. This results in uneven light emission along the track. The atmospheric muons typically consist of hundreds of muons and hence their light emission averages out, resulting in a more even distribution. This is shown in Fig. 2b.

3.2 Data set

The input data samples comprise of *events*, each consisting of the timings and the total charge deposited on each DOM. Each raw input image has the dimensions 86×60 laid out on a hexagonal grid, and is mapped onto an orthogonal grid to create an image of size $10 \times 20 \times 60$. More details about the dataset can be found in [1]. This study did not use information from the Deep Core DOMs, which are specialized DOMs placed near the center of the IceCube grid.

In [1], a few of the authors in this paper had explored the potential of Graph Neural networks (GNNs) and CNNs in performing signal-background classification. Their results

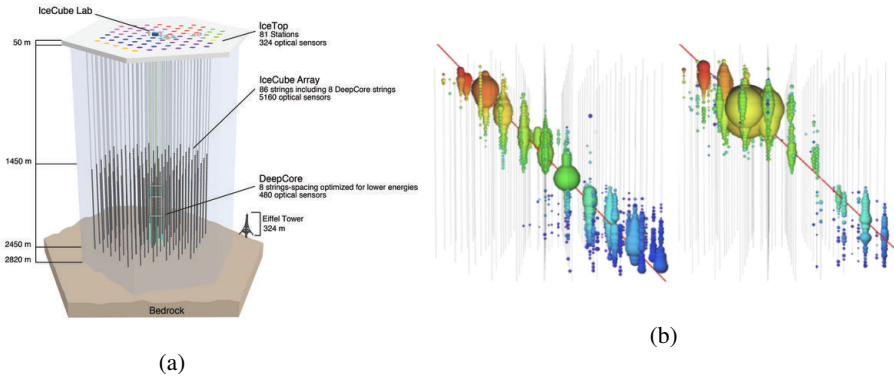


Figure 2: Fig (a) shows the setup of the IceCube Observatory. There are approximately 5000 Digital Optical modules (DOMs) placed within the Antarctic ice. Fig (b) shows the pattern of light deposition for signal and background. The colored bubbles indicate the relative time of arrival of light, with red being earliest and blue being the latest. The size of the bubbles is proportional to the number of photons. To the left, the pattern is more uniformly distributed, which is indicative of a background event produced by atmospheric muons. The pattern in the right figure has a large stochastic energy deposition which is indicative of a signal due to a muon produced from an astrophysical neutrino.

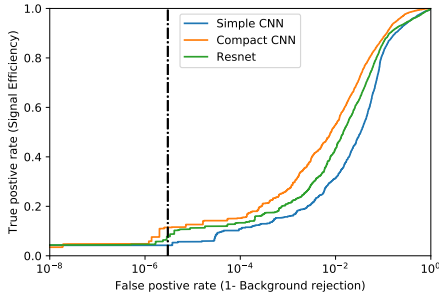
showed that both GNNs and CNNs performed better than the physics benchmarks, with the GNNs achieving better performance than the CNNs used in the paper (ResNet). In this work, we perform a more thorough exploration of CNN architectures. Another aspect that differentiates this work is the exclusion of a set of events called High Energy Starting Events (HESE) [7], which are neutrino events where the interaction starts inside the detector, and can be identified in a pre-selection stage using existing IceCube analyses.

For training and validation, we used 130989 samples, with a validation ratio of 33% and signal to background ratio of 16.2%. The test dataset had 737715 samples with a signal to background ratio of 1.92%.

3.3 Analysis and Results

The study in [1] utilized the *ResNet* CNN (similar to the ones in [8]), with very high number of parameters. In this analysis, we did a more thorough architecture search, specifically looking for compact, layered 3D CNNs with lesser number of parameters. Scanning CNN network architectures using combinations of convolution, pooling and dropout layers, we identified a few networks that achieved better performance than the previous ResNet model.

A good way to assess the performance of models is by looking at the Receiver Operating Characteristic (ROC) curves. Fig. 3a shows the ROC curves for three models: ResNet, the best CNN in this work termed *Compact CNN* and a CNN with very few parameters termed *Simple CNN* and Table 1 gives a comparison of these models. Comparing the true positive rate (tpr) values for the three models at a false positive rate (fpr) value of 3×10^{-6} (which was the value used for comparison with the physics cuts in [1]), it is clear that the *Compact CNN* performs better than ResNet, while having only a tenth of the parameters. Its structure is shown in Fig. 3b. Comparing the training times per epoch for the models in Table 1, it can be seen that the training time for the *Compact CNN* model is slightly higher than ResNet. Given that it has only a tenth of the parameters, this is a bit surprising. It is possible that



(a) Roc curves for 3 chosen models. The X axis denotes the false positive rate(fpr) and the Y axis denotes the True positive rate(tpr). The dotted line denotes an $fpr = 3 \times 10^{-6}$. The model *Compact CNN* from this work achieves best performance everywhere. We also compare these with a simpler model with very few parameters termed *Simple CNN*

Layer (type)	Output Shape	Param #
input_4 (InputLayer)	(None, 10, 20, 60, 1)	0
conv3d_9 (Conv3D)	(None, 10, 20, 60, 40)	3880
batch_normalization (BatchNo (None, 10, 20, 60, 40)		160
max_pooling3d_9 (MaxPooling3 (None, 10, 10, 20, 40)		0
dropout_7 (Dropout)	(None, 10, 10, 20, 40)	0
conv3d_10 (Conv3D)	(None, 10, 10, 20, 40)	153640
batch_normalization_1 (Batch (None, 10, 10, 20, 40)		160
max_pooling3d_10 (MaxPooling (None, 10, 5, 6, 40)		0
dropout_8 (Dropout)	(None, 10, 5, 6, 40)	0
conv3d_11 (Conv3D)	(None, 10, 5, 6, 40)	153640
batch_normalization_2 (Batch (None, 10, 5, 6, 40)		160
max_pooling3d_11 (MaxPooling (None, 10, 2, 2, 40)		0
dropout_9 (Dropout)	(None, 10, 2, 2, 40)	0
conv3d_12 (Conv3D)	(None, 10, 2, 2, 80)	307280
conv3d_13 (Conv3D)	(None, 10, 2, 2, 120)	921720
flatten_3 (Flatten)	(None, 4800)	0
dense_6 (Dense)	(None, 120)	576120
dense_7 (Dense)	(None, 1)	121
Total params: 2,116,881		

(b) Structure of the best performing model. It has ~ 2 million parameters.

Figure 3

Table 1: Comparison of selected trained models. tpr and fpr stand for true positive rate and false positive rate respectively, while AUC stands for area under the ROC curve. *Simple CNN* is a layered CNN with very few number of parameters, *Compact CNN* is the best performing model and ResNet is the model used in the previous paper [1]. The training times are obtained by running on a Titan X (Pascal) GPU.

Model type	# of parameters	Training time (s / epoch)	AUC score	Tpr at $fpr=3 \times 10^{-6}$
Simple CNN	14,789	23	0.922	0.04
Compact CNN	2,116,881	120	0.960	0.116
ResNet	26,486, 126	115	0.935	0.078

ResNet might be faster due to the connectivity between its different layers that is absent for our models. Nevertheless, we have developed a CNN with better classification performance than ResNet, while reducing the parameters by an order of magnitude.

4 Dataset 2: ATLAS

4.1 ATLAS experiment

The ATLAS experiment is one of the major experiments at the Large Hadron Collider (LHC) at CERN, Switzerland. It was one of the two LHC experiments involved in the discovery of the Higgs boson in 2012. Among other things, one of the main goals of the LHC is to look for evidence for Physics beyond the Standard Model of particle physics, such as *Supersymmetry*. In this work, we use a dataset that is an input for analyses searching for new

Table 2: Comparison of selected trained models. *tpr* and *fpr* stand for true positive rate and false positive rate respectively. *Old CNN* denotes the CNN used in the previous work, *Compact CNN* is the best performing CNN in this work. We also compare these with a ResNet model. The training times are obtained by running on a Titan X (Pascal) GPU.

Model type	# of parameters	Training time (s / epoch)	tpr at $fpr=3\times 10^{-3}$
Old CNN	34,515,201	294	0.641
ResNet	23,597,826	515	-
Compact CNN	43,009	40	0.746

massive supersymmetric (RPV-Susy) particles in multi-jet final states [9]. The goal is to use CNNs directly on low-level detector data from the entire calorimeter, without reconstruction of jets or tuning of analysis variables.

4.2 Data set

The dataset consists of simulated data obtained using the *Pythia* event generator [10] interfaced to the *Delphes fast detector simulation* [11, 12]. The Signal events are the RPV-Susy events, while the background is QCD. The images are 2D, with dimensions 64 x 64, with each image pixel representing the energy deposited in the calorimeter. For training and validation and testing, we used 412416, 137471 and 137471 samples respectively, with a signal to background ratio of about 43%.

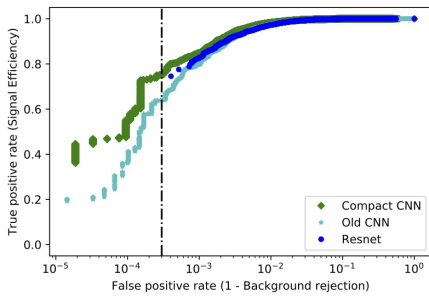
4.3 Analysis and Results

In [2], the potential of CNNs for signal-background classification was explored with this dataset. A few simple CNNs with large number of parameters were studied and they were found to achieve better performance than the physics benchmarks. The best model was found to have around 34 million parameters. The main aim of this work was to perform an architecture search to identify CNNs with better performance and fewer parameters.

As in the previous case, we explored network architectures using a combination of convolution, pooling and dropout layers. Table 2 and Fig. 4a show a comparison of the best performing model termed *Compact CNN* with the model used in the previous work and ResNet. From the ROC curves for these 3 models in Fig. 4a, it is clear that Compact CNN performs better than the other two models. It is also much simpler in structure (almost 1/800th the number of parameters) compared to the other two models, while having a significantly lower training time per epoch as seen in Table 2. The structural details of the Compact CNN model is given in Fig. 4b. Thus, we have developed a substantially more compact network, while reducing the number of parameters by almost 3 orders of magnitude.

5 Summary

We have demonstrated the effectiveness of compact, layered Convolutional Neural Networks in classifying signal from background for simulated datasets from two different particle physics experiments. In both cases, performing an architecture search, we have identified compact neural networks that achieve better performance than previous studies with substantially fewer parameters. Given this success of CNNs in signal-background classification, there is potential for their applicability to classification problems in other fields beyond particle physics.



(a) Roc curves for 3 chosen models. The X axis denotes the false positive rate(fpr) and the Y axis denotes the True positive rate(tpr). The dotted line denotes an $fpr = 3 \times 10^{-6}$. The Compact CNN achieves best performance everywhere, excelling especially in the low fpr region to the left.

Model: "model"

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 64, 64, 1)]	0
conv2d (Conv2D)	(None, 64, 64, 10)	100
max_pooling2d (MaxPooling2D)	(None, 32, 32, 10)	0
dropout (Dropout)	(None, 32, 32, 10)	0
conv2d_1 (Conv2D)	(None, 32, 32, 10)	910
max_pooling2d_1 (MaxPooling2D)	(None, 16, 16, 10)	0
dropout_1 (Dropout)	(None, 16, 16, 10)	0
conv2d_2 (Conv2D)	(None, 16, 16, 10)	910
max_pooling2d_2 (MaxPooling2D)	(None, 8, 8, 10)	0
dropout_2 (Dropout)	(None, 8, 8, 10)	0
flatten (Flatten)	(None, 640)	0
dropout_3 (Dropout)	(None, 640)	0
dense (Dense)	(None, 64)	41024
dense_1 (Dense)	(None, 1)	65

Total params: 43,009
 Trainable params: 43,009
 Non-trainable params: 0

(b) Structure of the best performing model. It has $\sim 43,000$ parameters.

Figure 4: Performance of compact layered CNNs.

6 Computational details

All computations were performed at NERSC. The CNNs were implemented in *keras* [13]. During the course of this work, we developed a package for training of general stacked CNN models with visualization tools¹. These should be of use for general purpose CNN training and visualization

7 Acknowledgments

We would like to thank the IceCube collaboration for providing access to the simulation dataset. We would also like to thank Spencer Klein and Nick Choma for useful discussions about the IceCube part of this work. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231. V.A.'s work was supported by the Computational Center for Excellence, a Computational HEP program in the Department of Energy's Science Office of High Energy Physics (Grant #KA2401022). S.R.'s work was supported by the National Science Foundation under grant number PHY-1307472.

References

- [1] Choma et al, *arxiv: 1809.06166*, 2018
- [2] Bhimji et. al, *arxiv: 1711.03573*, 2017
- [3] Ciresan, Meier, Schmidhuber, *arxiv: 1202. 2745*, 2012.
- [4] LeCunn, Bottou, Bengio, Haffner, *Proc of the IEEE, November 1998*.

¹https://github.com/vpayyar/layered_CNNs_for_ATLAS_data

-
- [5] IceCube collaboration, *The Astrophysical Journal*, Volume 833, Number 1, 2016.
 - [6] IceCube collaboration, *Journal of Instrumentation*, Volume 12, Issue 03, pp.P03012, 2017.
 - [7] IceCube collaboration, *Science* Vol 342, Issue 6161, 1242856, 2013. doi:10.1126/science.1242856.
 - [8] K. He, X. Zhang, S. Ren, and J. Sun, *arXiv:1512.03385*, *Proc. CVPR*, 2016.
 - [9] ATLAS Collaboration 2016 Search for massive supersymmetric particles in multi-jet final states produced in pp collisions at $s = 13$ TeV using the ATLAS detector at the LHC Tech. Rep. , *ATLAS-CONF-2016-057 CERN Geneva* ,<http://cds.cern.ch/record/2206149>
 - [10] Sjostrand T, Mrenna S and Skands P, 2008 *Computer Physics Communications* 178 852 – 867, 2008
 - [11] de Favereau J, Delaere C et al. 2014 , *JHEP* 2014 57
 - [12] Cacciari M, Salam G P and Soyez G 2012, FastJet, *The European Physical Journal C* 72 1896
 - [13] Chollet, F. , keras, GitHub. <https://github.com/fchollet/keras>, 2015