Deep Learning based Algorithms in Astroparticle Physics

Martin Erdmann, Jonas Glombitza
RWTH Aachen University, III. Physikalisches Institut A, Otto-Blumenthal-Str., 52056 Aachen, Germany

E-mail: jonas.glombitza@physik.rwth-aachen.de

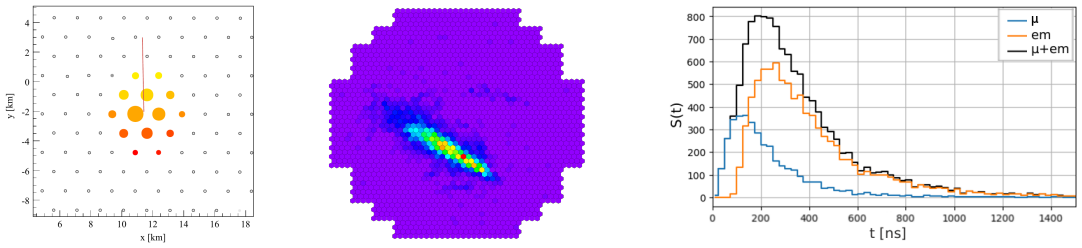
Abstract. In recent years, great progress has been made in the fields of machine translation, image classification and speech recognition by using deep neural networks and associated techniques (deep learning). Recently, the astroparticle physics community successfully adapted supervised learning algorithms for a wide range of tasks including background rejection, object reconstruction, track segmentation and the denoising of signals. Additionally, the first approaches towards fast simulations and simulation refinement indicate the huge potential of unsupervised learning for astroparticle physics. We summarize the latest results, discuss the algorithms and challenges and further illustrate the opportunities for the astrophysics community offered by deep learning based algorithms.

1. Introduction

The latest deep learning based algorithms have shown to be extraordinary successful across many domains in computer vision, engineering and science [1, 2, 3, 4]. In physics research, especially for experiments located at the LHC, various of these recent algorithms have already been investigated. In particle physics the concept of deep networks and supervised learning has been used for signal and background separation in a wide range of task. In jet physics these techniques were used for identifying the underlying partons [5, 6] or measuring the substructure of jets [7]. Furthermore, first attempts towards reinforcement learning were proposed at this conference [8, 9]. Additionally, deep learning based algorithms were used for object reconstruction, including the reconstruction of tracks [10] and particle energies using calorimeters [11].

Very recently, generative adversarial networks [12] (GANs) have shown great progress in the field of representation learning and generative modelling. These most recent approaches were found to significantly speed up physics simulations [13, 14] for the upcoming high-luminosity LHC, where simulations are required on a much larger scale than before.

In general, many challenging tasks in particle physics are similar to these in astroparticle physics. Similarly, a major challenge is the separation of a small number of signal events from large backgrounds, for example in gravitational wave [15], gamma [16] and neutrino astronomy or for the search of ultra-high energy photons [17]. In addition to background rejection and event classification, the reconstruction of objects or important physics observables from huge amounts of data is a typical task. Furthermore, the measured signals often contain noise, which needs to be suppressed to ensure good reconstructions. These extremely challenging tasks involving large amounts of data motivate the use of modern methods like deep learning in astroparticle physics.



(a) Air shower footprint measured at the Pierre Auger Observatory [18].

(b) Hexagonal grid of PMT's as used by H.E.S.S. [19].

(c) Simulated signal trace of a water-Cherenkov station [20].

Figure 1: Typical datastructure of astrophysical measurements

This work is structured as follows. First, the typical structure of measured data of astrophysics experiments is introduced. Afterwards, several successful applications of supervised learning based on convolutional and recurrent neural networks are discussed and compared to traditional reconstruction algorithms. Finally, first unsupervised deep learning applications are reviewed and interpreted with a view to the future.

2. Data

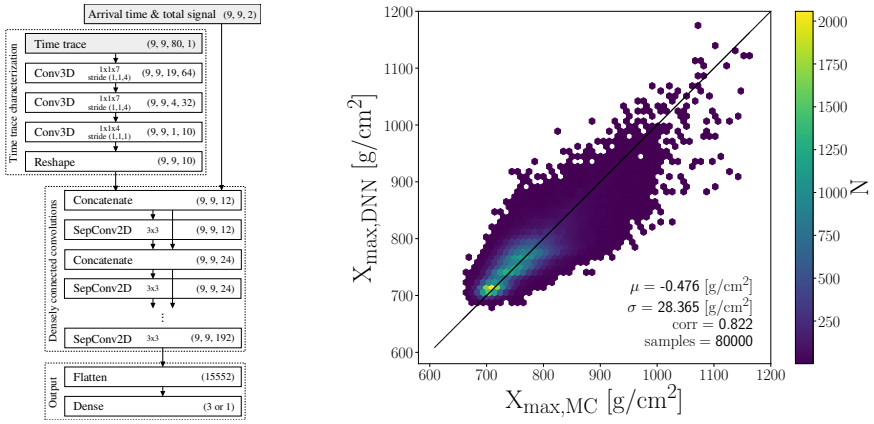
In elementary particle physics, many deep learning algorithms are based on the processing of the 4-momenta of measured particles in the respective particle detector. Often several traditional cleaning, clustering or selection algorithms are already applied in early stages of the analysis chain to preprocess the data for the deep neural networks. In contrast to the mostly compact detectors with high readout density as used in elementary particle physics, typical astrophysics experiments feature much larger detector volumes and a sparse readout. The sparse readout is mostly caused by cost reasons and limited computational resources at the sites, because many experiments are located in very isolated places like the Argentinian pampa or Antarctica. The sensors of these large detectors are often placed on hexagonal or triangular grids. To still allow for convolutions, which are normally defined on cubic grids, axial or offset coordinate systems are commonly used.

Beside the spatial structure of the detector placement, the signal measurements often contain exact time information. This degree of symmetry in the data motivates the use of convolutional and recurrent network architectures. Convolutional and recurrent architectures exploit the symmetry of data to reduce model parameters substantially. This greatly simplifies the numerical problem to be solved, hence stabilizes the network training and improves the reconstruction performance.

3. Supervised Learning Applications

3.1. Air Shower Reconstruction at the Pierre Auger Observatory

Ultra-high energy cosmic ray (UHECR) observatories like the Pierre Auger Observatory [18] measure the energy spectrum of cosmic rays, study their composition and search for their origins. Due to their very high energies ($E > 10^{18}\ eV$), cosmic rays induce large particle cascades, when penetrating the Earth's atmosphere. These particle cascades interact further with the atmosphere and create footprints at the Earth's surface with sizes of several 10 km^2 which are measured by surface-detector stations. A key to search for the UHECR sources is the cosmic-ray charge, which is needed to correct for deflections of the galactic magnetic field. To obtain an estimate of the cosmic-ray charge, the atmospheric depth of the shower maximum $X_{\max}$ is a suitable observable. It decreases with increasing mass due to a growing number of



(a) (b)

Figure 2: (a) Network architecture for air shower reconstruction. (b) Resulting performance for the reconstruction of the depth of shower maximum.

sub showers. Due to the indirect measurement and the large fluctuations in the air-shower development, the reconstruction of the primary cosmic-ray charge is a challenging task.

3.1.1. Neural Network Architecture and Training. In the sample setup presented here, the measurement of a cosmic ray induced air shower contains a signal trace (cfg. 1) with 80 timesteps and timing information measured by each surface detector station, leading to 3-dimensional cube of signal traces and 2-dimensional time map. Since all the simulated air showers fit in a window of 9×9 stations, a region centered to the station with the highest signal is used as a selection cut. Additionally to the timing information, a map holding the total signals of the respective stations is used as additional feature map.

The fundamental network architecture proposed by [20] consists of 3 parts. In the first part, the measured signal traces are processed using 3 convolutional layers. The used convolution only takes place in the time dimension of the signal, hence the filters processing the shape of the signal trace are shared over the stations. This stabilizes the training process and ensures that features do not mix spatially in the first part of the network (Fig. 2 a). In the beginning of the second part, the additional feature maps are concatenated to the output of the first part. Here, a block of densely connected separable convolutions processes the air-shower footprint, by extracting and combining features in space and time. Afterwards, a final output layer predicts the respective depth of the shower maximum. All in all the model holds $\approx 90,000$ parameters.

3.1.2. Results. The network is trained on $\approx 360,000$ air showers. Due to the relatively small model, no overtraining can be observed. Further, the increase in model size and the use of regularisation could not boost the performance. For the training showers with zenith angles between $\theta \in [0°, 60°]$ and energies between $1 - 100$ EeV following a spectrum of E^{-1} are used. The showers are simulated with a parametrized air-shower simulation and a corresponding detector simulation inspired by [21, 22]. The composition consists of proton, helium, nitrogen and iron nuclei in equal fractions. In figure 2 b the resulting correlation between the deep learning based reconstruction and the true simulated depth of shower maximums is shown. In general the high correlation coefficient indicates a good performance. Furthermore, the distribution of reconstructed events is unbiased $\mu \approx 0$ and shows an excellent resolution of $\sigma \approx 28\text{g/cm}^2$.

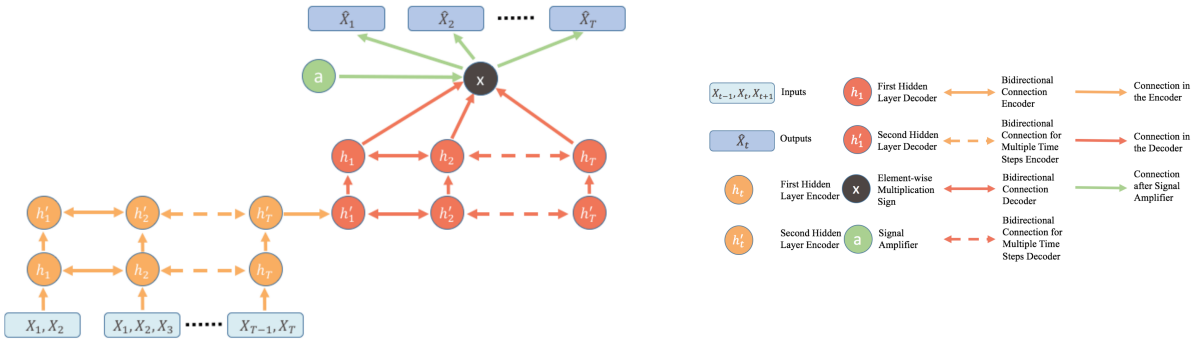


Figure 3: Network architecture of recurrent denoising autoencoder for denoising gravitational waves [23].

3.2. Denoising of Gravitational Waves at LIGO

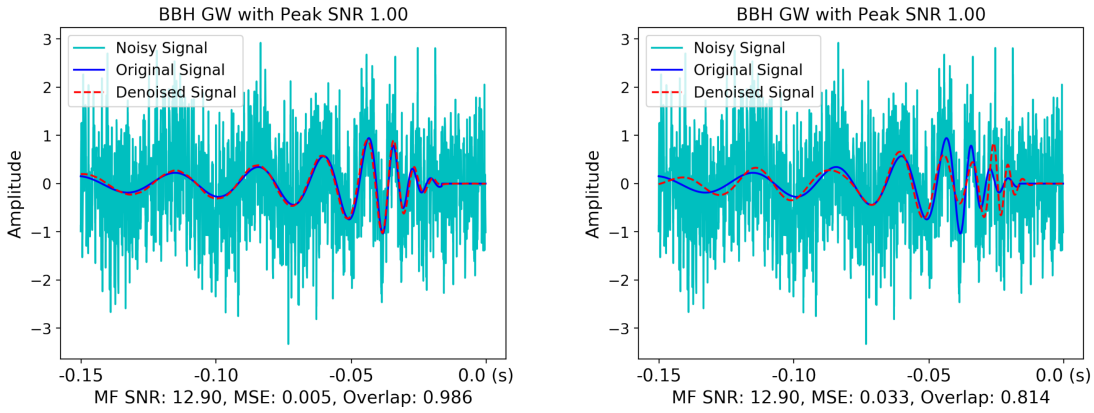
After their recent discovery, gravitational-wave astronomy rapidly evolved into an important research field of astrophysics. Nowadays, several detector sites measure disturbances in the curvature of spacetime using large Michelson interferometers. The detection of binary black-hole mergers is most difficult due to the high contribution of noise in the gravitational-wave measurements. Hence, it is of major importance to remove that noise and reconstruct the pure signal waveform.

Recently a denoising autoencoder (DAE) to reconstruct the gravitational-wave signal at observatories like LIGO was introduced [23]. The typical task for DAEs is to restore a specific sample x given a corrupted sample $\tilde{x}$ using a function $f(\tilde{x})$ which is usually represented by a neural network. Due to the time structure of the gravitational-wave measurements the usage of recurrent sequence-to-sequence models is beneficial.

3.2.1. Neural Network Architecture and Training. Like a typical autoencoder, the recurrent DAE consists of two parts: an encoder and a decoder. The encoder features 2 long short-term memory [24] (LSTM) layers with 64 units each. As input, it receives the time step to denoise and additionally 8 neighbor time steps. Hence, the denoising is not only dependent on a single time step, but includes the temporal structure of the measurement. Furthermore, using *bidirectional* LSTM layers helps to better resolve correlations between earlier and later time steps in the traces. The output of the encoder is the final hidden-state vector, which serves as input for the decoder which holds again 2 bidirectional LSTM layers with 64 units. The decoder part reconstructs the signal from the state vector and outputs the respective denoised time step.

3.2.2. Results. The recurrent DAE is trained on a dataset of simulated gravitational waveforms that describe binary black-hole mergers with mass ratios < 10 in steps of 0.1 and total masses between 5 and 75 solar-masses in steps of 1. The waveforms are generated with a sampling rate of $8,192\ kHz$ and whitened with the sensitivity of LIGO. Furthermore, Gaussian noise and random time shifts up to 15% of the total trace length are added. To stabilize the training, a form of *curriculum learning* is used, which slowly increases the complexity of the task by decreasing the signal-to-noise ratio (SNR) during the training. In addition, the changing the SNR is a form of data augmentation that reduces the risk of overtraining.

In figure 4 a denoised waveform obtained using the deep learning approach is compared to a model using 10 principal components on a test sample. The DAE shows a much better overlap with the underlying signal trace and overall a much better performance than the Principal Components Analysis (PCA), especially for late stages of the signal. Note that, for testing,



(a) (b)

Figure 4: Denoised gravitational waveform using (a) recurrent denoising autoencoder (DAE) and (b) PCA model using 10 components. [23]

non-Gaussian noise measured by LIGO around a gravitational-wave candidate was injected to the traces.

4. Unsupervised Learning Applications

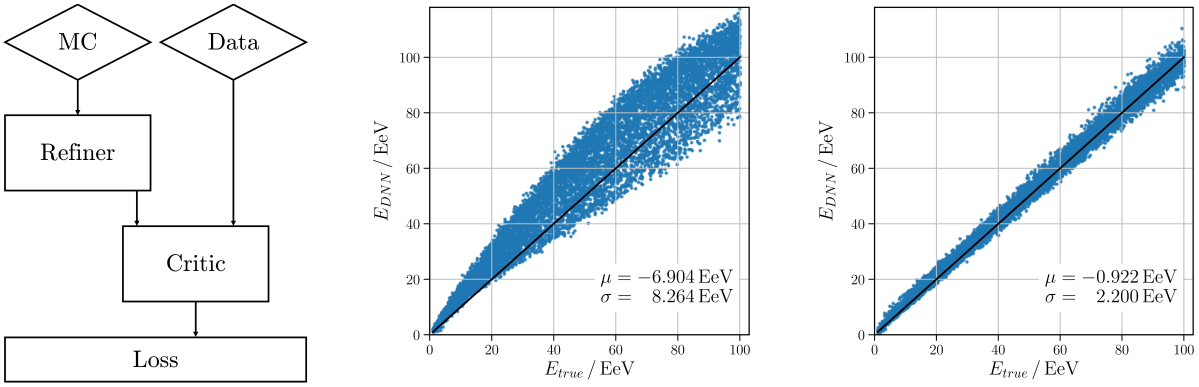
Besides supervised learning algorithms, also unsupervised applications in the scope of astroparticle physics were published. Due to the recent progress in the field of generative adversarial networks (GANs), several applications incorporating adversarial frameworks were utilized. In this context, the first applications targeting on the challenge of fast simulations were investigated. Here, the focus of the underlying task is mostly related to density estimation and to study similarity of probability densities.

Most machine learning algorithms in physics are trained on extensive Monte Carlo simulations and later applied to measured data. Therefore, the generalization error when facing the trained model measured data depends very much on the quality of the simulation. Mismatches between data and simulation, such as multidimensional correlations will lead to an inductive bias which distorts the model performance on measured data. Refining simulations, to be more similar to data is one approach to reduce this systematic bias.

4.1. Framework for Simulation Refinement

The proposed refiner framework [25] looks similar to GANs, but incorporates decisive differences. The fundamental task of the framework is to train a refiner which receives simulated events as input and outputs refined simulations which better resemble measured data. The measure of the refiner performance is given by a critic network, similar as in Wasserstein GANs [26]. Here, the critic estimates the Wasserstein distance between refined and measured events. To allow for a reasonable estimation, the gradient penalty is used in the critic as proposed in [27] and trained 10 iterations before the refiner is updated once. Beside providing a meaningful distance measure, approximating a smooth metric in the critic network like Wasserstein-1 improves the stability of the framework due to non-vanishing and meaningful gradients. By minimizing the Wasserstein distance between the refined and measured samples during the refiner updates, the quality of the refined simulations increases during the iterations.

To enforce the refiner network to induce only small changes of the simulation, the architecture of the refiner networks consists of 3 residual blocks holding 2 convolutional layers each. The



(a) (b) (c)

Figure 5: (a) Architecture of the framework, used to refine simulations. Resulting performance when evaluating a network trained on (b) the original simulation and (c) refined simulation on data [25].

architecture of the critic network is similar to the network for air shower reconstruction and is based on densely connected separable convolutions.

4.1.1. Refining Detector Traces. The framework is trained for 10,000 refiner iterations using two datasets: *simulation* and "*data*" which is also simulated but in a different way. Both datasets contain 100,000 air showers with energies between $1 - 100$ EeV following E^{-1} and zenith angles between $\theta \in [0°, 60°]$. The difference between the datasets is the fraction between the electromagnetic and the muonic component in the air showers. The simulation features a split of 70/30, whereas the (simulated) data features a split of 30/70, and an increased noise level. These differences should illustrate a mismodeling in the simulation. Note, that for an application using real measured data, a second measurement is needed which is ensured in most air-shower observatories due to a hybrid calibration with fluorescence telescopes.

As input for the refiner, signal traces of measured air showers are used. In this specific set up the refiner should learn to correct the wrong component fraction encoded in the simulation in a semi-supervised manner, by comparing simulated and real traces. To evaluate the performance gained using the refined simulation, another neural network is trained to reconstruct the cosmic-ray energy using the refined traces and evaluated on data. This is compared to the generalization performance when using a model trained on the original simulation.

Figure 5 shows that the energy reconstruction of the network trained on the refined simulation shows a much better performance on data. The bias is reduced and the resolution has improved substantially. In contrast, the network trained on the original simulation, shows large errors due to the mismodeling of the component fraction, which is especially visible at the highest energies. To some extent, the refiner was able to correct for the deviating component fraction, which improved the reconstruction performance.

5. Conclusion and Outlook

Recently, the astroparticle community successfully adapted machine learning algorithms to typical tasks like signal reconstruction and processing. Due to regular sensor placements in astrophysics experiments which incorporate symmetries in space and time, recurrent and convolutional architectures show the most promising results. The spread of successful applications includes a variety of tasks like reconstruction, denoising and classification. Basically,

most concepts can be applied to various experiments as the respective model can be adapted to capture the respective symmetry of the detector design.

The first example discussed object reconstruction in the context of cosmic-ray physics. The approach of using deep neural networks for reconstructing physics observables using the detected raw signals showed promising results. Besides the reconstruction of physics observables, for many astroparticle physics experiments the preparation and denoising of raw signals is a major challenge. The presented method for the denoising of gravitational-wave signals using recurrent denoising autoencoders illustrates an application which addresses these challenges. In addition, the first steps for reducing the inductive bias due to differences between simulations and data were discussed. Using an adversarial framework the original simulation was refined to better resemble the data. The results indicate the potential of deep learning based algorithms beyond pure classification and regression.

Further investigating the systematics of machine learning models in order to control and better understand predictions is a key task for the future. The strength of the presented algorithms relies on their speed and precise reconstruction when analyzing huge amounts of data. All these are advantages for future experiments which will be faced with even larger amounts of data, which have to be processed and interpreted on shorter time scales.

Acknowledgments

This work is supported by the Ministry of Innovation, Science and Research of the State of North Rhine-Westphalia, and the Federal Ministry of Education and Research (BMBF).

References

[1] Yu D and Deng L 2014 *Automatic Speech Recognition: A Deep Learning Approach.* (Springer)
[2] He K, Zhang X, Ren S and Sun J 2015 *CoRR* (*Preprint* 1512.03385)
[3] Ronneberger O, Fischer P and Brox T 2015 *CoRR* (*Preprint* 1505.04597)
[4] Goodfellow I, Bengio Y and Courville A 2016 *Deep Learning* (MIT Press) www.deeplearningbook.org
[5] Baldi P, Sadowski P and Whiteson D 2015 *Phys. Rev. Lett.* **114** 111801 (*Preprint* 1410.3469)
[6] Erdmann M, Fischer B and Rieger M 2017 *JINST* **12** P08020 (*Preprint* 1706.01117)
[7] Baldi P, Bauer K, Eng C, Sadowski P and Whiteson D 2016 *Phys. Rev.* **D93** 094034 (*Preprint* 1603.09349)
[8] Carrazza S and Dreyer F A 2019 *Phys. Rev.* **D100** 014014 (*Preprint* 1903.09644)
[9] Erdmann M, Fischer B and Noll D (eds) 2019 *Reinforced sorting networks for particle physics analyses* to appear in the ACAT 2019 proceedings
[10] Farrell S and et al (eds) 2017 *Particle track Reconstruction with Deep Learning*
[11] Carminati F, Khattak G and Pierini M (eds) 2017 *Calorimetry with Deep Learning*
[12] Goodfellow I J, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A and Bengio Y 2014 Generative adversarial networks (*Preprint* 1406.2661)
[13] Paganini M, de Oliveira L and Nachman B 2018 *Phys. Rev. Lett.* **120** 042003 (*Preprint* 1705.02355)
[14] Erdmann M, Glombitza J and Quast T 2019 *Comput. Softw. Big Sci.* **3** 4 (*Preprint* 1807.01954)
[15] Adams T S, Meacher D, Clark J, Sutton P J, Jones G and Minot A 2013 *Phys. Rev.* **D88** 062006
[16] Ohm S, van Eldik C and Egberts K 2009 *Astroparticle Physics* **31** 383391 ISSN 0927-6505
[17] Aab A *et al.* (Pierre Auger) 2017 *JCAP* **1704** 009 (*Preprint* 1612.01517)
[18] Aab A *et al.* (Pierre Auger) 2015 *Nucl. Instrum. Meth.* **A798** 172–213 (*Preprint* 1502.01323)
[19] The HESS Collaboration, 2019, Largest ever cherenkov telescope sees first light, www.mpi-hd.mpg.de/hfm/HESS/pages/press/2012/HESS_II_first_light/
[20] Erdmann M, Glombitza J and Walz D 2018 *Astropart. Phys.* **97** 46–53 (*Preprint* 1708.00647)
[21] Heck D, Knapp J, Capdevielle J N, Schatz G and Thouw T 1998
[22] Ave M, Roth M and Schulz A 2017 *Astropart. Phys.* **88** 46–59
[23] Shen H, George D, Huerta E A and Zhao Z 2017 (*Preprint* 1711.09919)
[24] Hochreiter S and Schmidhuber J 1997 *Neural Computation* **9** 1735–1780
[25] Erdmann M, Geiger L, Glombitza J and Schmidt D 2018 *Comput. Softw. Big Sci.* **2** 4 (*Preprint* 1802.03325)
[26] Arjovsky M, Chintala S and Bottou L 2017 Wasserstein gan (*Preprint* 1701.07875)
[27] Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V and Courville A 2017 Improved training of wasserstein gans (*Preprint* 1704.00028)