

Reconstruction Techniques in IceCube using Convolutional and Generative Neural Networks

Mirco Huennefeld^{1,*} for the IceCube Collaboration

¹TU Dortmund University, Dortmund, Germany

Abstract. Reliable and accurate reconstruction methods are vital to the success of high-energy physics experiments such as IceCube. Machine learning based techniques, in particular deep neural networks, can provide a viable alternative to maximum-likelihood methods. However, most common neural network architectures were developed for other domains such as image recognition. While these methods can enhance the reconstruction performance in IceCube, there is much potential for tailored techniques. In the typical physics use-case, many symmetries, invariances and prior knowledge exist in the data, which are not fully exploited by current network architectures. Novel and specialized deep learning based reconstruction techniques are desired which can leverage the physics potential of experiments like IceCube.

A reconstruction method using convolutional neural networks is presented which can significantly increase the reconstruction accuracy while greatly reducing the runtime in comparison to standard reconstruction methods in IceCube. In addition, first results are discussed for future developments based on generative neural networks.

1 Reconstruction Techniques in IceCube

A key challenge to the success of experiments such as IceCube is the reliable and accurate reconstruction of events. In IceCube, further challenges arise as the detector is situated at the geographic South Pole where resources are limited. However, to perform real-time analyses and to issue alerts to telescopes around the world, powerful reconstruction methods are desired. This results in a dilemma as performance is often paired with computational complexity. But even for offline reconstructions, the computational complexity of the most advanced maximum-likelihood methods can render these intractable and hence limit the physics potential.

Machine learning based methods might help to alleviate these complications. Deep neural networks can be extremely powerful and their usage is computationally inexpensive once the networks are trained. These characteristics make an approach based on deep learning an excellent candidate for application in IceCube.

Convolutional neural networks (CNNs) [1] have been previously applied in neutrino experiments [2] and can also greatly enhance the reconstruction performance in IceCube as shown in [3]. However, convolutional architectures have considerable limitations. Novel deep learning based methods specifically tailored to the needs in high-energy physics experiments such as IceCube are needed.

*e-mail: mirco.huennefeld@tu-dortmund.de

2 Event Reconstruction with Convolutional Neural Networks

The CNN based reconstruction method as presented in [3] is further improved and extended to also include the reconstruction of cascade events in IceCube. A detailed description of the enhanced reconstruction method will follow in a future publication.

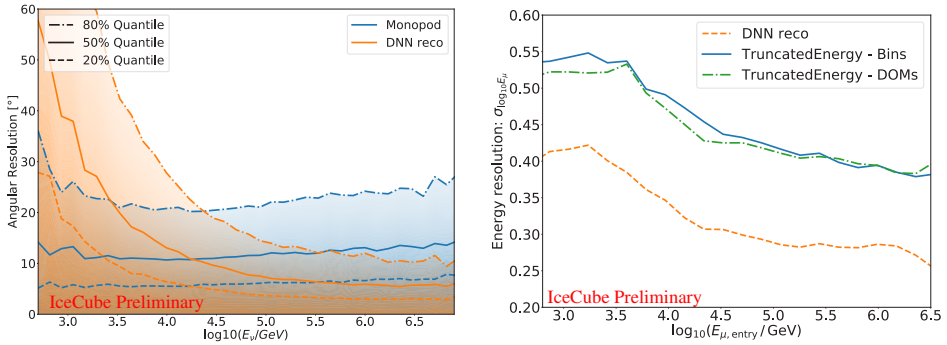


Figure 1. On the left, the angular resolution for cascade events is shown for IceCube’s current best reconstruction method (Monopod [4]) and for the newly developed method based on deep neural networks (DNN reco). On the right, the muon energy resolution is shown for the current standard methods (variations of TruncatedEnergy [4, 5]) and the deep learning based method.

The performance of the CNN based reconstruction method is compared to the current standard in Fig. 1. With the help of the neural network reconstruction, the muon energy resolution can be greatly improved over the whole energy range. In addition, the angular resolution of cascade-like events can be increased by almost a factor of two at higher neutrino energies, while reducing the runtime by 2-3 orders of magnitude. This is of considerable impact, because the high-energy events are more likely to be of astrophysical origin. The worse resolution at lower energies is due to limited statistics in the training sample.

Though originally designed for an efficient and fast application on-site at the South Pole, the developed CNN method is able to compete with the most sophisticated and advanced offline reconstruction methods in IceCube.

3 Limitations of Convolutional Neural Networks

Despite its success, the CNN method presented above has considerable limitations. The network architecture assumes translational invariance and the data to be aligned on a regular grid. However, these assumptions are only approximately fulfilled in IceCube. Most importantly though, prior knowledge cannot easily be exploited by the network architecture as it could, for instance, in maximum-likelihood methods. The use of other standard architectures, such as graph neural networks (GNNs) as applied in [6], can help with the irregularities in the IceCube detector grid, but generally face the same challenges when trying to include additional knowledge.

A complementary approach is to develop novel techniques to exploit prior knowledge and to combine strengths of maximum-likelihood and deep learning methods. The big breakthrough in the domain of image recognition is based on convolutional neural networks and the exploitation of symmetries and a priori knowledge [1, 7]. Convolutional architectures can utilize translational invariance and locality to greatly reduce the number of free parameters.

In the physics use-case, many more symmetries, invariances, and physical laws exist which are yet to be exploited. An approach how this could be accomplished is described in the following section for the reconstruction of cascade-like events in IceCube.

4 Cascade Reconstruction with Generative Networks

For a maximum-likelihood method, one generally compares measured data to some expectation value. This can, for instance, be the number of measured photons at a specific photo-multiplier. The expected number of photons is then obtained from Monte Carlo simulations. This process, while being the most accurate, is extremely slow and hence not always feasible. Reconstruction of a single event may take hours to days in such an approach.

Instead of performing computing intensive simulations, a generative network can be used to approximate them. A neural network, the generator, is trained to predict the expected waveforms at each digital optical module (DOM) for a given cascade hypothesis. The full cascade hypothesis is defined by seven parameters: (x, y, z, t , azimuth, zenith, energy). Once the generator network is trained, it can then be used in reverse mode. The waveforms for a given event can now be compared to the generated waveforms for a given hypothesis. A distance measure is computed between the generated and true waveforms. This distance (e.g., negative log likelihood or $\Delta\chi^2$) is a function of the cascade hypothesis and can be minimized analogous to a maximum-likelihood approach.

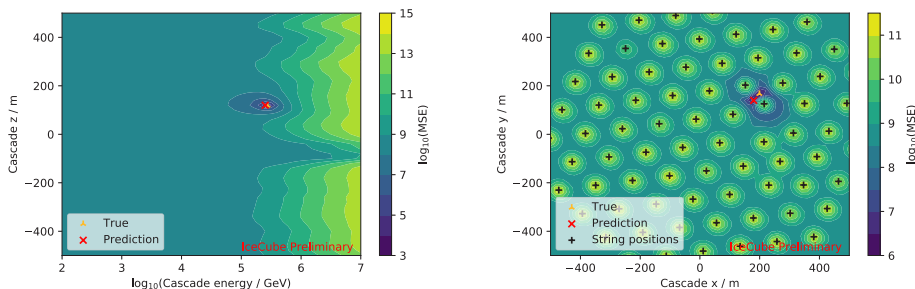


Figure 2. The parameter space of the cascade hypothesis is visualized by performing a scan in two variables, while keeping the others fixed to their true values. A scan of the cascade energy vs. the z -coordinate of the vertex is shown on the left. On the right, a scan in x - and y -coordinates is displayed. Plots are adapted from [8].

An idealized example is shown in Fig. 2. Two of the seven cascade parameters are kept free, while the others are fixed to their true values. For every point in the landscape, the generator is used to obtain the expected waveforms which are then compared to the true waveforms. A distance measure, in this case the mean squared error (MSE), is calculated and plotted to obtain a scan of the landscape. As shown in Fig. 2, the true values can be recovered to a reasonable resolution for this example event.

Similarly, the cascade could have been resimulated for every point in the landscape and then compared to the true event. Such a simulation would require many CPU and GPU hours, while the generator network (once trained) only requires a few seconds. In contrast to the simulation, the generator network is fully differentiable, allowing gradient descent methods to be used to minimize the distance measure. In this case, the per-event reconstruction time

with the generator network further reduces to $\mathcal{O}(100 \text{ ms})$ as opposed to the brute force scan that takes $\mathcal{O}(s)$.

Although the generator is only given relative coordinates of the DOMs to the cascade vertex during training, it is able to learn the geometry of the IceCube detector as shown on the right of Fig. 2 where the IceCube string positions are overlaid. In addition, the generator can identify the dust layer at depths between about $0 < z < -150 \text{ m}$. At these depths, the glacial ice at the South Pole exhibits a broad peak in dust concentration [9]. As a result, cascades in the dust layer will seem dimmer than they actually are. This relation is also evident in the MSE landscape shown on the left of Fig. 2.

As opposed to the CNN method, the generator approach can utilize translational invariance in the cascade hypothesis rather than the approximative translational invariance in measured data which is affected by the detector acceptance and ice inhomogeneities. Moreover, irregularities in the detector grid are now naturally accounted for and prior knowledge such as how the waveforms are parameterized can easily be included in the generator architecture to reduce the number of free parameters.

5 Conclusion and Outlook

Standard neural network architectures such as CNNs can be utilized to further improve the event reconstruction accuracy, while greatly reducing the per-event runtime. They are well suited for the application online at the South Pole. Despite their success, these methods have considerable limitations. High-energy physics experiments typically have developed detailed simulations for their detectors. Hence, extensive knowledge exists about the data generation process, constraints, and physics laws the data obeys. Yet, current deep learning architectures can not fully exploit this information. Novel deep learning based reconstruction techniques, specifically tailored to the needs in high-energy physics experiments such as IceCube, are therefore desired which can combine the strengths of existing architectures and maximum-likelihood methods. The presented approach using generative neural networks provides a promising step towards these goals.

References

- [1] Y. LeCun et al., *Adv Neural Inf Process Syst* **2**, 396 (1990)
- [2] A. Aurisano et al., *JINST* **11**, P09001 (2016)
- [3] M. Huennefeld (IceCube Collaboration), *PoS ICRC2017*, 1057 (2017)
- [4] M. Aartsen et al. (IceCube Collaboration), *JINST* **9**, P03009 (2014)
- [5] R. Abbasi et al. (IceCube Collaboration), *Nucl. Instrum. Meth.* **A703**, 190 (2013)
- [6] N. Choma et al., arXiv:1809.06166 [cs.LG] (2018)
- [7] G. Hinton et al., *Adv Neural Inf Process Syst* **25**, 1097 (2012)
- [8] A. Flores and M. Huennefeld, SFB876-C3, TU Dortmund, Report No. **5**, (2018)
- [9] M. Ackermann et al., *J. Geophys. Res. Atmos.* **111**, D13203 (2006)