

Air-Shower Reconstruction at the Pierre Auger Observatory based on Deep Learning

Jonas Glombitza^{*†} for the Pierre Auger Collaboration^{†b}

^a*RWTH Aachen University, Aachen, Germany*

^b*Observatorio Pierre Auger, Av. San Martín Norte 304, 5613 Malargüe, Argentina*

E-mail: auger_spokespersons@fnal.gov

Full author list: http://www.auger.org/archive/authors_icrc_2019.html

The surface detector array of the Pierre Auger Observatory measures the footprint of air showers induced by ultra-high energy cosmic rays. The reconstruction of event-by-event information sensitive to the cosmic-ray mass, is a challenging task and so far mainly based on fluorescence detector observations with their duty cycle of $\approx 15\%$. Recently, great progress has been made in multiple fields of machine learning using deep neural networks and associated techniques. Applying these new techniques to air-shower physics opens up possibilities for improved reconstruction, including an estimation of the cosmic-ray composition. In this contribution, we show that deep convolutional neural networks can be used for air-shower reconstruction, using surface-detector data. The focus of the machine-learning algorithm is to reconstruct depths of shower maximum. In contrast to traditional reconstruction methods, the algorithm learns to extract the essential information from the signal and arrival-time distributions of the secondary particles. We present the neural-network architecture, describe the training, and assess the performance using simulated air showers.

36th International Cosmic Ray Conference — ICRC2019

24 July – 1 August, 2019

Madison, Wisconsin, USA

^{*}Speaker.

[†]for collaboration list see PoS(ICRC2019)1177

1. Introduction

The search for the origin of ultra-high energy cosmic rays (UHECRs) is one of the greatest challenges of astroparticle physics. Studying the mass of these particles helps to localize possible source candidates and to understand the acceleration mechanisms at the sources. UHECRs exceed primary energies of 10^{18} eV leading to extensive air showers when impacting the Earth's atmosphere. To allow for accurate measurements of UHECRs, the Pierre Auger Observatory [1] was completed in 2008 featuring a hybrid design. The baseline of the Observatory is the surface detector [2] (SD) formed by 1660 water-Cherenkov stations placed in a hexagonal grid, which measure the footprint of extensive air showers induced by UHECRs. The stations are separated by a distance of 1500 m covering an area of ≈ 3000 km². This surface detector grid is overlooked by 27 fluorescence telescopes which measure the light emission of nitrogen excited by the particles of the extensive cascade.

The air-shower development is especially dependent on the energy and the mass of the primary cosmic ray, causing greater depth of shower maximum for lighter masses and higher energies. Hence, the atmospheric depth of the shower maximum X_{\max} enables an estimation of the cosmic-ray composition [3]. The observable X_{\max} can be directly measured with the fluorescence detector (FD). However, measurements of the FD are only possible on moonless nights, which reduces the duty cycle to $\approx 15\%$. Alternatively, measuring the shower maximum using the SD is a challenging task, since it has to be inferred indirectly from the secondary particles recorded on the ground.

Recently, great progress has been made in machine learning using deep neural networks and associated techniques (deep learning) [4, 5]. The basic elements of deep learning are neural networks holding adaptive weights, which can be adjusted to the task at hand by exploring a data set. The use of graphics processing units (GPUs), large amounts of data and the latest developments have recently made it possible to train deep networks which hold several million parameters. Especially in computer vision [6], pattern [7] and speech [8] recognition, using deep networks paved the way for results of previously unreachable precision [4, 9].

The symmetric placement of sensors in a regular grid is ideally suited for the use of deep-learning techniques. Besides using pattern recognition for analyzing the spatial characteristic of the footprint, the arrival-time distribution of the secondary particles is encoded in the time evolution of the measured signal (signal trace) in each water-Cherenkov station. Recognizing single muons and particles of the electromagnetic component to infer the longitudinal shower development further encourages the use of techniques of speech recognition or natural language processing.

Apart from the total signal and the signal start time as used in the standard reconstruction [10], the machine-learning approach relies on the time-dependent distributions of particles traversing the water-Cherenkov stations. In contrast to shower-universality analyses [11, 12] where humans parametrize the signal and arrival-time distributions of secondary particles using analytic models, the deep-learning algorithm discovers the best parametrization itself.

In this contribution we show that using deep networks enables the reconstruction of several air-shower properties by combining spatial and time information of the detected air shower. First, we introduce the deep-convolutional network architecture and discuss the training. Subsequently, we show that the network is capable of reconstructing the basic shower geometry and providing a reconstruction of the primary particle energy. Finally, we show that the deep-convolutional neural

network provides promising results in the reconstruction of the depth of the shower maximum X_{\max} on an event-by-event level.

2. Surface detector data

The air-shower footprint triggers several water-Cherenkov stations, thus inducing a characteristic signal pattern on the hexagonal SD grid. Each of the triggered stations provides a signal trace which encodes a time-dependent density of particles traversing the station. The resulting data are illustrated in Figure 1.

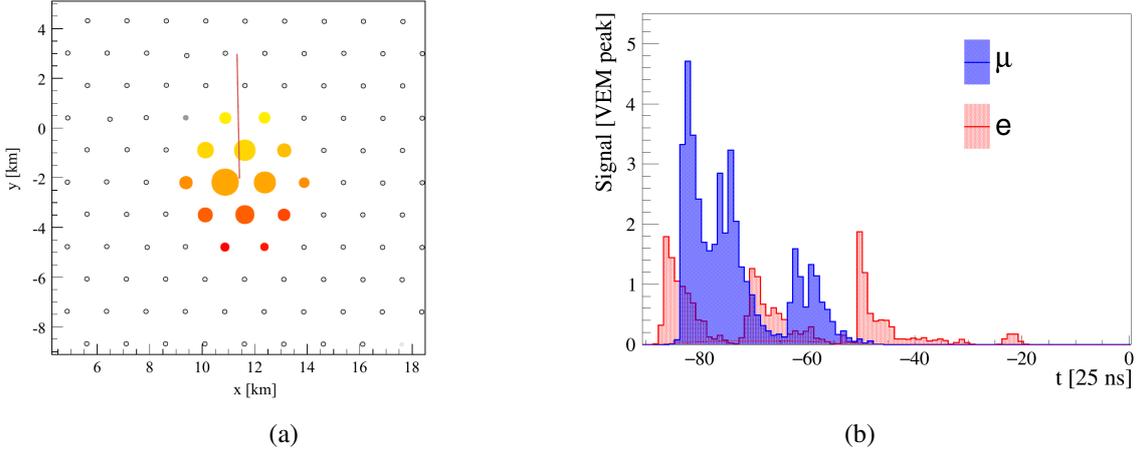


Figure 1: (a) Measured surface-detector signal pattern, induced by a simulated air shower. (b) Simulated time traces measured at one surface-detector station.

2.1 Preprocessing

In order to stabilize the training of the deep network, we used data preprocessing. By normalizing the input variables to $[1, -1]$, we observed better training stability and faster convergence. To further enforce all mappings in the neural network to act on the same scale, we adopted a standard normalization ($\sigma = 1$, $\mu = 0$) of our label distributions.

2.1.1 Signal patterns and coordinates

Due to the hexagonal grid of the surface detector, we cannot make direct use of convolutions which rely on Cartesian representations. In order to use convolutions, we transform our signal patterns using an axial representation, by padding with zeros and shifting each second row until alignment is achieved [13]. As a measured footprint usually has a size of several tens of square kilometers, the pattern is cropped to a window of 13×13 stations to reduce memory consumption. The station with the largest signal is used as the center of this frame and defines the origin of the local coordinate system.

As the placement of the stations shows slight deviations from a perfect grid, we use the coordinates of each station as additional inputs. We first transform the x and y coordinate of each event, into the respective local coordinate system and calculate the deviation with respect to a perfect

grid. Additionally, we normalize the deviation by dividing by σ_C , which is the spread of the distribution of grid deviations. For the altitude z of the respective station we use standard normalization ($\sigma_z = 1$, $\mu_z = 0$), otherwise the relative height would be calibrated out, which would cause shifts of the observed atmospheric depth. In order to use the same representations, the position of the shower core is transformed in the same way.

2.1.2 Signal traces

As the number of particles measured in each detector is approximately exponentially distributed, we use a logarithmic re-scaling of the trace:

$$\tilde{S}_i(t) = \frac{\log_{10}[S_i(t) + 1]}{\log_{10}[S_{sat} + 1]}. \quad (2.1)$$

To allow for positive values only we use $S_i(t) + 1$. The signal amplitude $S_i(t)$ is linearly re-scaled into the range $[0; 1]$ using the minimum and maximum signal level S_{sat} that can be digitized by the simulated data-acquisition system.

2.1.3 Arrival time

The arrival direction of the primary particle can be reconstructed using the arrival times $t_{0,i}$ of the shower front at the stations. For each station, the arrival time is normalized with respect to the arrival time τ_{center} measured at the central station and $\sigma_{r, \text{data set}}$, the standard deviation of the arrival time distribution:

$$\tilde{t}_{0,i} = \frac{t_{0,i} - \tau_{center}}{\sigma_{r, \text{train data set}}}. \quad (2.2)$$

2.1.4 Station states

Due to aging effects, bad calibrations or broken PMTs, some of the water-Cherenkov stations may not be ready to measure real data. In order to let the network correct for this, we add a feature map encoding the station states (1 = ready to measure, 0 = broken) as additional input. During the training, we further imitate broken stations by randomly masking stations and propagating this information to the feature map of stations states.

3. Deep network for air shower reconstruction

The network design for reconstructing cosmic-ray induced air showers features three stages and is based on [13]. The first stage is used for the characterization of the time trace, hence learning features which include information about the shower development. The following part allows forming spatial features. In the final stage, several sub-parts for each reconstruction task are used.

3.1 Time trace characterization

The first part relies on characterizing the measured signal trace. Investigating the time-dependent signals allows for measuring the content of the muon and electromagnetic components of the air shower which each induce different signal shapes. The separation of the components yields information about the shower age, energy and the primary mass. To let the network exploit local

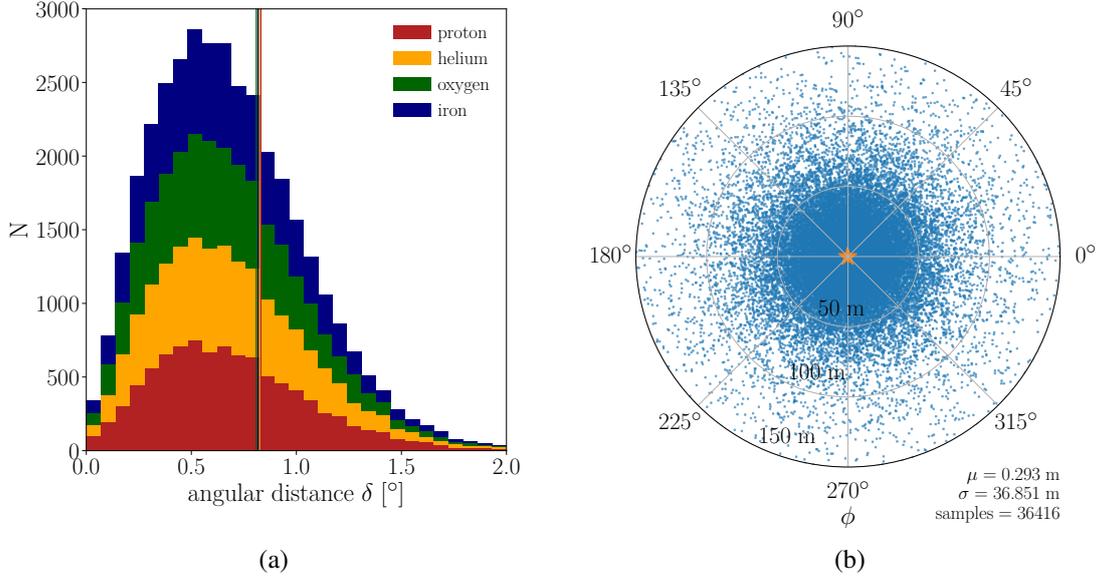


Figure 2: (a) Angular distance between the reconstructed and the true shower axes. Different colors indicate different primaries. (b) Azimuth dependence of the shower-core reconstruction. The network provides on average an unbiased estimate of the core position (yellow marker).

features in the signal traces, e.g., sharp edges or spikes, we apply convolutions. We use rectified linear units [5] as activation functions and three convolutional layers with varying kernel sizes. To stabilize the training and to increase generalization performance, we use weight sharing along the spatial dimensions. Hence, the convolutional operation works on the time dimension only and reduces the time trace with a length of 251 time steps, to 10 features. In each station the same type of features are extracted which allows the concatenation of station-dependent inputs in the following part.

3.2 Spatial representation

After the characterization of the signal traces, the extracted features are used as input for the second part and concatenated with the additional feature maps holding the arrival times, station coordinates and the station states. The following network structure is based on densely connected [14] and separable convolutions [15]. This stabilizes the training process and provides important physics parameters like the arrival time in each layer. We observed that using batch normalization adds too much disturbance in the training. Hence, as activation functions we used scaled exponential linear units [16] to normalize the data between the layers.

3.3 Task-related part

The final part of the neural-network architecture consists of individual network parts for each reconstruction task. This last stage is based on residual units [6] and pooling operations. In total, we used two residual blocks covering three residual units each. Each of the residual units contains two layers using a rectified linear unit in between. As the last layer we used global average pooling.

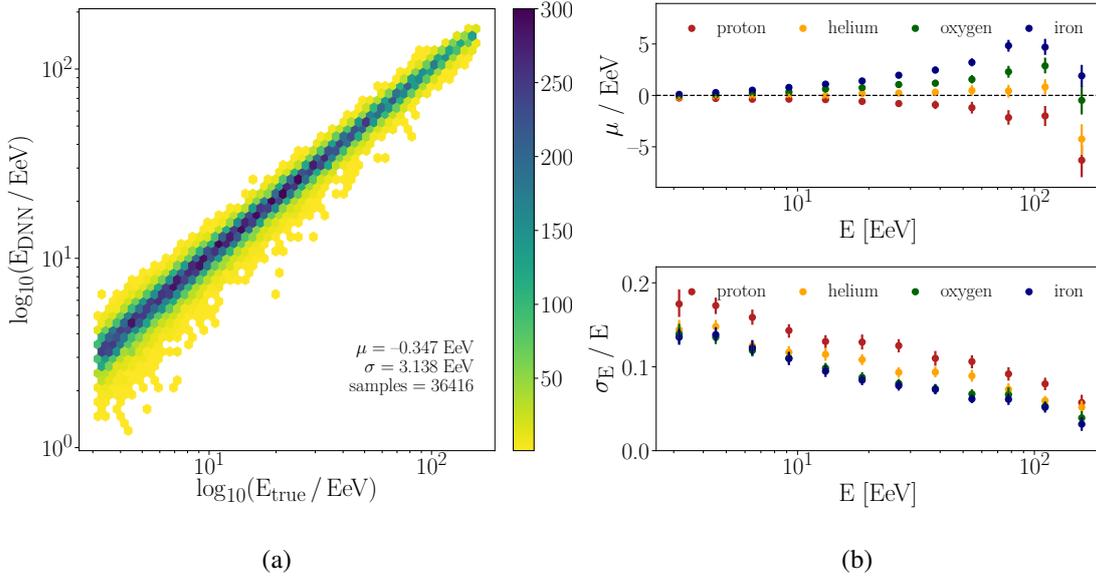


Figure 3: (a) Correlation between reconstructed and true energies. (b) Energy-dependent bias and resolution of the reconstruction shown for different primaries.

4. Performance in simulations

We used in total $\approx 400,000$ air showers simulated with CORSIKA [17] using the hadronic interaction model EPOS-LHC [18]. The data is simulated in an energy range of 1 – 160 EeV following a spectrum of E^{-1} . As primary particles, protons, helium, oxygen and iron nuclei are used covering a full azimuth range and a zenith range of $\theta \in [0^\circ, 65^\circ]$. We used 360,000 showers for training the network, 4,000 for the validation and around 37,000 for the final evaluation.

4.1 Training

We trained the model for 27 epochs on a Nvidia GeForce GTX 1080 graphics card which took ≈ 7 h with a batch size of 32. For training the network, we used Keras [19] with the TensorFlow [20] backend. As optimizer Adam [21] was used with its default settings. We further smoothly decayed the learning rate and dropped the learning rate by a factor of 1.33 when the validation loss did not decrease after three epochs. As loss for all tasks, we used the mean squared error and re-weighted the losses to the same order of magnitude.

4.2 Geometry reconstruction

Before investigating the reconstruction of complex shower variables like the shower maximum, the network needs to be capable of providing a reconstruction of the basic air-shower properties. For the validation of the deep-learning based reconstruction, we first investigated the quality of the reconstructed shower geometry. In Figure 2a, the angular distance between the reconstructed and the true shower axis is illustrated, showing a good angular resolution (68% quantile) of $q_{68} = 0.8^\circ$. The colors represent different primary particles. Figure 2b shows the reconstruction

of the shower core. The absolute resolution amounts to $\sigma \approx 37$ m. Further, no azimuth-dependent bias is visible which implies that the network is able to correct for shower asymmetry effects.

4.3 Reconstruction of the energy

After validating the basic reconstruction of the shower geometry, the reconstruction of the primary cosmic-ray energy can be studied. Figure 3a shows the reconstruction of the energy. The overall reconstruction shows no bias. In Figure 3b the different resolutions and biases are plotted for different primaries. The composition bias depends on the energy and increases for higher energies. The relative energy resolution improves for higher energies due to decreasing fluctuations and achieves a relative resolution of $\sigma_E/E \leq 10\%$ at around 100 EeV.

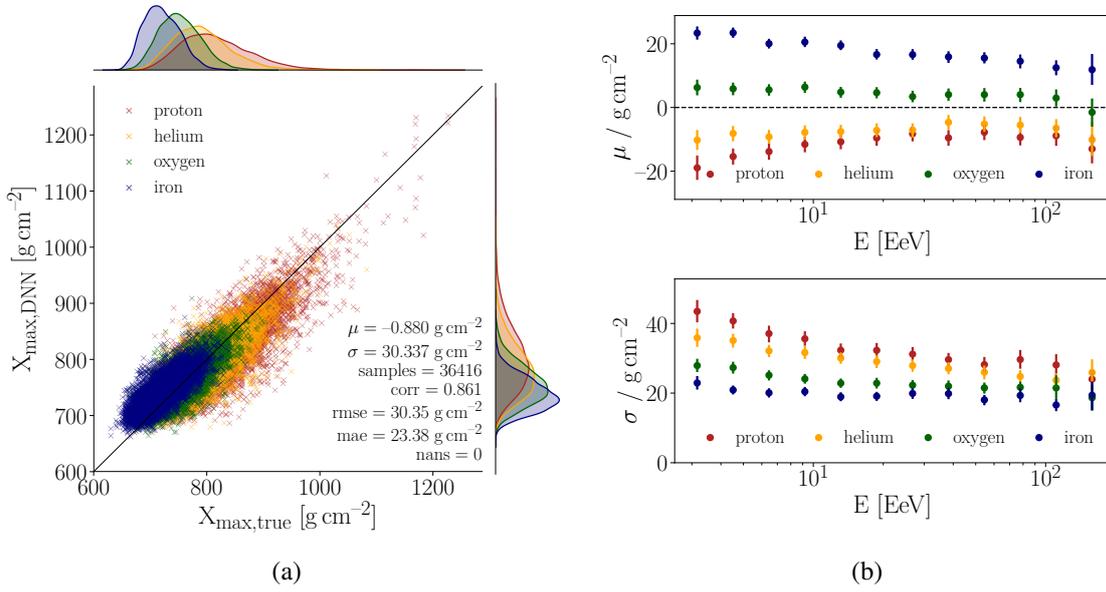


Figure 4: (a) Correlation between reconstructed and true depths of shower maximum X_{\max} . (b) Energy-dependent resolution and bias of the X_{\max} reconstruction.

4.4 Reconstruction of the shower maximum

With the precision described previously, the network is able to reconstruct the air-shower geometry and estimate the primary cosmic-ray energy. In the following the reconstruction of the shower maximum X_{\max} is investigated. In Figure 4a, the reconstructed shower maxima are shown for proton, helium, oxygen and iron. The correlation between the true X_{\max} and the reconstruction amounts to 0.8. It can be seen that the elements show separation in the reconstructed shower maxima, and hence the algorithm can be used for an estimation of the cosmic-ray mass. The overall resolution of the reconstruction is $\approx 30 \text{ g/cm}^2$. The energy dependence of the resolution and bias is shown in Figure 4b for different primaries. The bias decreases from $\approx 40 \text{ g/cm}^2$ at 3 EeV to $\leq 25 \text{ g/cm}^2$ at 150 EeV. Due to the decreased shower-to-shower fluctuations, the reconstruction of heavier elements shows an improved resolution of $\sigma_{X_{\max}} \approx 25 \text{ g/cm}^2$ at the highest energies.

5. Conclusion

Extensive air showers induce time-dependent signal patterns at the surface detector of the Pierre Auger Observatory. To reconstruct the properties of the primary cosmic ray we used a deep neural network. The presented algorithm relies on convolutions exploring huge amounts of data. We have discussed the intrinsic structure of our model, which learns time-dependent representations in a first stage. In the second stage the network allows for the extraction and combination of features in space and time. We have demonstrated that deep learning can be used to reconstruct the geometry and energy of the air shower using simulated surface-detector data. In addition, the reconstruction of the shower maximum X_{\max} showed very promising results. Further validating the performance with fluorescence detector measurements, and investigating different hadronic interaction models, would allow for a mass-composition estimation with the full SD event statistics.

References

- [1] A. Aab et al. [Pierre Auger Collaboration], The Pierre Auger Cosmic Ray Observatory, *Nucl. Instrum. Meth. A* **798** (2015) 172.
- [2] A. Aab et al. [Pierre Auger Collaboration], The Surface Detector System of the Pierre Auger Observatory, *Nucl. Instrum. Meth. A* **586** (2008) 409.
- [3] A. Aab et al. [Pierre Auger Collaboration], Depth of Maximum of Air-Shower Profiles at the Pierre Auger Observatory: Composition Implications, *Phys. Rev. D.*, **12** (2014) 1.
- [4] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, Cambridge, MA, US (2016).
- [5] Y. LeCun, Y. Bengio, G. Hinton, Deep Learning, *Nature* **521** (2015) 436.
- [6] K. He et al., Deep residual learning for image recognition, [arXiv/1512.03385](https://arxiv.org/abs/1512.03385).
- [7] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, [arXiv/1505.04597](https://arxiv.org/abs/1505.04597).
- [8] D. Yu, L. Deng, Automatic Speech Recognition: A Deep Learning Approach, Springer, London, UK (2014).
- [9] O. Russakovsky et al., Imagenet large scale visual recognition challenge, *IJCV* **115** (2015) 211.
- [10] S. Argiro et al., The Offline software framework of the Pierre Auger Observatory, *Nucl. Instrum. Meth. A* **580** (2007) 1485.
- [11] A. M. Hillas, Angular and energy distributions of charged particles in the electron photon cascades in air, *J. Phys. G.* **8** (1982) 1461.
- [12] M. Ave, M. Roth, A. Schulz, A generalized description of the time dependent signals in extensive air shower detectors and its applications, *Astropart. Phys.*, **88** (2017) 46.
- [13] M. Erdmann, J. Glombitza, D. Walz, A Deep Learning-based Reconstruction of Cosmic Ray-induced Air Showers. *Astropart. Phys.* **97** (2018) 46.
- [14] G. Huang et al., Densely Connected Convolutional Networks, [arXiv/1608.06993](https://arxiv.org/abs/1608.06993).
- [15] F. Chollet, Xception: Deep Learning with Depthwise Separable Convolutions, [arXiv/1610.02357](https://arxiv.org/abs/1610.02357).
- [16] G. Klambauer, T. Unterthiner, A. Mayr, S. Hochreiter, Self-Normalizing Neural Networks, [arXiv/1706.02515](https://arxiv.org/abs/1706.02515).
- [17] D. Heck et al., CORSIKA: A Monte Carlo code to simulate extensive air showers, Forschungszentrum Karlsruhe Report FZKA (1998) 6019.
- [18] T. Pierog et al., EPOS LHC : test of collective hadronization with LHC data, [arXiv/1306.0121](https://arxiv.org/abs/1306.0121).
- [19] F. Chollet, Keras: Deep Learning for Python (2015), github.com/fchollet/keras.
- [20] M. Abadi et al., TensorFlow: Large-scale machine learning on heterogeneous systems (2015), [tensorflow.org](https://www.tensorflow.org).
- [21] D. P. Kingma, J. Ba, Adam: A Method for Stochastic Optimization, Proc. 3rd Int. Conf. on Learning Representations (ICLR), San Diego (2015), [arXiv/1412.6980](https://arxiv.org/abs/1412.6980).