PROCEEDINGS OF SCIENCE

Cosmic-Ray Composition analysis at IceCube using Graph Neural Networks

Paras Koundal [a,*] **for the IceCube collaboration** [†]

[a] *Institute for Astroparticle Physics, Karlsruhe Institute of Technology, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Karlsruhe, Germany*

E-mail: paras.koundal@kit.edu

The IceCube Neutrino Observatory is a multi-component detector embedded deep within the South-Pole Ice. This proceeding will discuss an analysis from an integrated operation of IceCube and its surface array, IceTop, to estimate cosmic-ray composition. The work will describe a novel graph neural network based approach for estimating the mass of primary cosmic rays, that takes advantage of signal-footprint information and reconstructed cosmic-ray air shower parameters. In addition, the work will also introduce new composition-sensitive parameters for improving the estimation of cosmic-ray composition, with the potential of improving our understanding of the high-energy muon content in cosmic-ray air showers.

**** 27th European Cosmic Ray Symposium - ECRS ****
**** 25-29 July 2022 ****
**** Nijmegen, the Netherlands ****

[†]Full Author List : https://icecube.wisc.edu/collaboration/authors/
[*]Speaker

PoS(ECRS)085

1. Introduction

IceCube is a cubic-kilometer astroparticle detector at the geographic South Pole. Over 5000 digital optical modules (DOMs) are deployed on 86 strings in glacial ice at depths ranging from 1450 m to 2450 m [1]. It detects particles from astrophysical sources, and tries to understand the dynamics of the sources through their cosmic ambassadors, allowing analysis in the areas of cosmic rays, neutrino physics, and other research areas. The in-ice IceCube (IC) array is accompanied by the surface component, called IceTop (IT) [2]. In addition to this IceCube also has a DeepCore (DC) array with DOM density roughly five times higher than that of the standard IC array, located around the center of the IC array at depths below 2100 m [3]. This makes it a unique three-dimensional multimessenger detector. IT is predominantly used as an extensive air-shower (EAS) detector, in addition to serving as a veto for neutrino detection. In an integrated approach, IT (primarily detects the electromagnetic component + GeV muons) and IC+DC (primarily detects TeV muons) can be used to reconstruct the direction, energy, and mass of penetrating particles of the incident cosmic ray (CR). An illustration of an example air shower incident on IT with also a footprint in IC is shown in Figure 1.

IceCube has already demonstrated the ability to improve our understanding of the energy and mass spectrum in the transition region from galactic to extragalactic cosmic rays [4]. This work extends ongoing analyses [5, 6], where the cleaned signal footprint is used. This has been possible because of ongoing advancements in the field of deep-learning. For the estimation of primary mass a graph neural network (GNN) based approach has been used. The Monte-Carlo (MC) simulations used for the analysis and training were generated using the CORSIKA air-shower generator [7]. The datasets for each primary type (p, He, O and Fe) were simulated in the energy range $5.0 \leq \log_{10}(\mathrm{E/GeV}) \leq 8.0$, using FLUKA [8] as the low-energy hadronic interaction model and SIBYLL 2.1 [9] as the high-energy interaction model. For this investigation, we only consider coincident IT+IC events utilizing quality cuts [4] that guarantee coincidence, with successful direction reconstruction and shower-core containment within the IT array. This work will also introduce two new composition-sensitive parameters.

2. New Composition-Sensitive Parameters

IceCube has performed multiple analyses, where several shower-features using simulations were probed [5, 6] and validated on real data [4]. $\mathbf{\log_{10}(S_{125}/VEM)}$, where S_{125} is the signal expectation at a perpendicular distance of 125 m from the shower axis at IT, was found to be a good energy-proxy [10]. The signal is in units of Vertical Equivalent Muon (VEM), which is the signal that would be deposited by a single muon vertically traveling through an entire IT tank. The charge deposits at IT are also used to reconstruct directional information like zenith and azimuth [4]. For mass-discrimination we had developed multiple useful parameters [4–6, 11, 12]. Among those, the ones used in this study are *a)* $\mathbf{\log_{10}(dE/dX_{1500\ m})}$ = fit-value of IC energy-loss profile at a slant-depth of 1500 m (details in [4]) *b)* **total stochastic energy** = total IC energy of high-energy local-stochastic deposits in an event (details in [5]). These parameters primarily probe energy-deposits by TeV muons in the detector. In the following we discuss ongoing efforts to develop newer composition sensitive parameters.

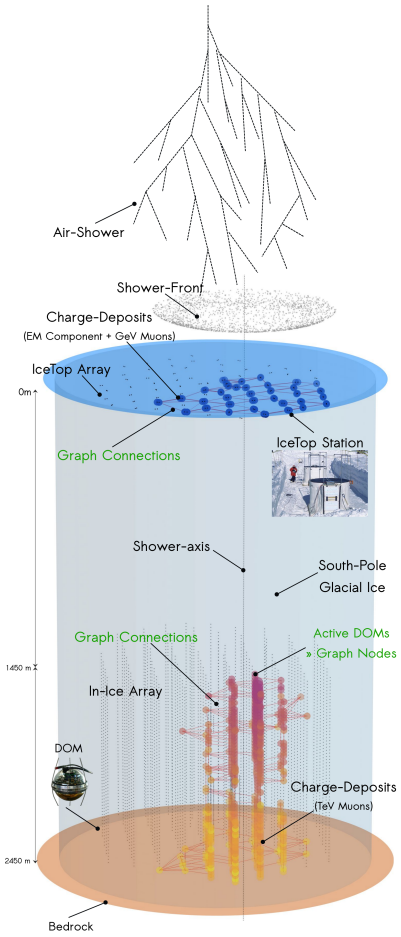

Figure 1: An illustration of an example EAS incident on IceTop. The dark-blue(/colored) circles represent the station(/DOM) hits after spatio-temporal cleaning around the reconstructed shower-axis. Only the hits are interpreted as nodes of a graph and the graph-connections are based on a predefined/adaptive algorithm.

2.1 Impact-Weighted Average Charge (IAC)

The information about the lateral-spread of in-ice muons was ignored in the previous studies. However, from our understanding of air-shower physics, at the same primary-energy Fe-induced air-showers are expected to have wider muon bundles than those initiated by p. This is anticipated because of aggregated behaviour of two observations from air-shower physics, namely at same energy: *a*) Fe initiated showers interact earlier in the atmosphere than p. *b*) Fe initiated showers have larger muon-multiplicity with lower average muon energy. For Fe showers, these phenomena allow for lower energy-muons with larger transversal momenta to be situated further away from the shower axis.

To test the hypothesis for IceCube, the IC spatio-temporal cleaned signal-footprint was divided along the slant-length into 5 percentile-bins (based on the number of DOM hits). The bins were divided based on percentiles (rather than depth) to ensure that we compare showers in the same zenith-bin (of the same energy and primary type) at approximately the same stage of shower development (independent of the point of first-interaction). This also prevents sparsity (and hence

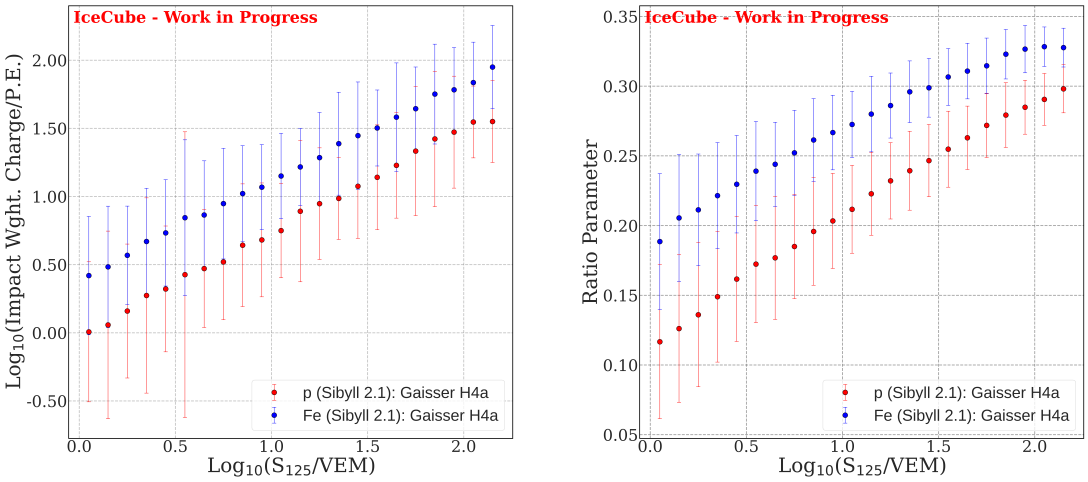

Figure 2: Composition sensitivity for new IceCube & IceTop variables (using SIBYLL 2.1 [9] and weighted to the H4a all-particle spectrum [13]): *Left:* Impact-Weighted Charge (last percentile bin), *Right:* Ratio Parameter; as a function of shower size, S_{125}, for proton and iron primaries.

over-reliance on few DOMs) for the following parameter calculation. In each percentile bin the Impact-Weighted Average Charge (IAC) is calculated, and represented as:

$$IAC_{Percentile\ Bin} = \frac{\sum_i C_i r_i}{\sum_i r_i} \tag{1}$$

where the summation runs over all the spatio-temporal cleaned hit-DOMs in a given percentile bin and C_i represents the charge at DOM i located at a perpendicular distance r_i (impact) from the reconstructed shower-axis. It was found that the last percentile bin shows the maximal composition sensitivity. The results for p-Fe separation, for the last percentile bin, is shown in Figure 2 (left). As can be seen, the parameter shows a promising discrimination power for CR composition analysis. Further studies to improve the discrimination power of the parameter are ongoing. A future analysis can also check the feasibility of using the parameter for non-coincident showers.

This study also explored the usage of Charge-Weighted Average Impact (CAI) as a mass discriminator. It also showed mass discrimination, though much lower than IAC. Hence, the plots are not shown here. For the GNN analysis (section 3), both IAC and CAI in all percentile bins were used.

2.2 Ratio Parameter

KASCADE-Grande was an air shower experiment located at the Karlsruhe Institute of Technology (average atmospheric depth 1022 g/cm^2) and has demonstrated that the ratio of muon-number to total charged-particle number is a good composition-sensitive parameter [14, 15]. IceCube lacks the capability of directly counting the particle-number for different shower-components. In order to test the feasibility of particle-type ratios for mass-discrimination at IceCube, reconstructed

proxy-parameters were used to approximate the muon-number (using $\mathbf{log_{10}(dE/dX_{1500\ m})}$) and total charged-particle number (using $\mathbf{log_{10}(S_{125})}$). The preliminary results from such a test are shown in Figure 2 (right). As can be seen from the figure, the parameter shows good discrimination power for CR composition analysis. This simple test also shows the potential benefits of obtaining particle-number estimates for composition analysis. Improvement in the mentioned parameter for IceCube and similar tests for IceCube-Gen2 are ongoing.

3. Cosmic-Ray Composition : Using Graph Neural Networks

The IceCube Collaboration plans to enhance the present detector to the next-generation instrument called IceCube-Gen2 [16]. The detector will enhance our understanding of the multimessenger universe by combining observations from multiple cosmic messengers. In addition to this, the enhancements will also extend the energy-sensitivity range and sky coverage of current measurements. The science objectives will be met by incorporating four new components: an in-ice optical array, a low-energy core, a surface air shower array [17], and an extended radio detector array. The enhancements will also shift the detector geometry to a more irregular one. The viability of utilizing GNNs for making composition analysis flexible to such detector upgrades has been demonstrated in prior studies [5, 6, 18]. This work extends that by aggregating older and newer composition-sensitive parameters, along with other relevant shower parameters. In addition to that, there have been technical changes in the GNN architecture, in comparison to [5]. The full network is trained as a regression-model with the logarithmic mass of CR primary i.e. ln(A) as the expected output.

3.1 Network Architecture

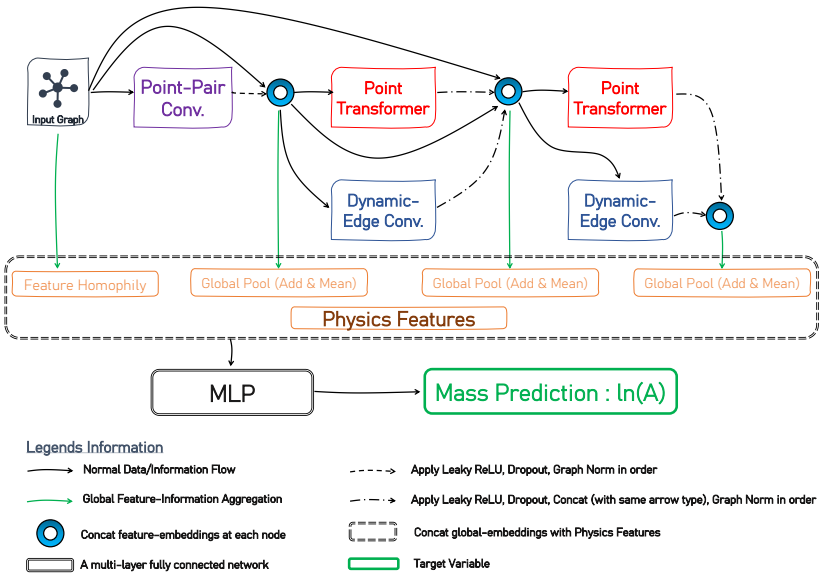

Figure 3: Current GNN based architecture used for CR analysis at IceCube Observatory.

The current detector can easily be mapped to an orthogonal geometry, which was the reason for an extensive use of convolutional neural network (CNN) based approaches at IceCube. However,

CNN for IceCube currently has a few minor problems like sparsity of DOM-hits (e.g. Figure 1) and treating DC [19–21] and IT [22] separately. These problems will be exaggerated even more as we move to IceCube-Gen2. GNNs are used in order to develop a method which allays these problems and is flexible to detector upgrades. We overcome the sparsity problem by mapping only the active stations/DOMs as nodes of a graph. The problem of combining IT, IC and DC is automatically solved by using GNNs, since they allow us to learn on irregular point clouds of variable size. Mapping IT as a graph and using it together with IC+DC is also a major addition in contrast to the previous study [5]. Additionally, the physics features included are *a*) $\log_{10}(S_{125})$ *b*) $\log_{10}(dE/dX_{1500m})$ *c*) total stochastic energy *d*) IAC and CAI (all percentile bins) *e*) Ratio Parameter *f*) track length (In-Ice) *g*) reconstructed Zenith and Azimuth .

To map the detector (IT+IC+DC) as a graph, the active stations/DOMs (hits remaining after signal cleaning) are mapped as nodes of the graph. Each node of the graph has associated attributes, which capture the spatial coordinates, the charge and timing information of the measured waveform for the hit DOMs. Defining node-neighbourhood i.e. edge-information at each node in the input-graph is another component essential for learning on graphs. Multiple methods were tested to define a neighbourhood. For the input-graph, **weighted k-NN with inductive-bias** has proven to be the best option. In our context: *a*) k-NN means connect each node to its k Nearest-Neighbors (in coordinate space) *b*) weighted k-NN means weight the k-NN connections between two nodes by spatial separation between them. *c*) inductive-bias means the node connections indirectly capture the information of graph (or event) size and the primary-type (by scaling Ratio Parameter described in subsection 2.2 - SRP). In summary, each node of an input-graph (or event) with N nodes has k neighbours, where:

$$k = 1 + \left\lceil \frac{N}{1 + A \cdot exp[SRP \cdot (B - B \cdot SRP \cdot N)]} \right\rceil \tag{2}$$

where A and B are two positive-valued hyperparameters of the network. The detailed reasons and tests for choosing the form of Equation 2 will be published in a future publication.

The steps described earlier give us an input graph (labelled as Input Graph in Figure 3), which can then be used as an input for the GNN architecture. Feature homophily gives a measure of how similar individual node-features of the connected-nodes are to each other [23]. The rest of the architecture is motivated from developments in powerful CNN architectures detailed in [24, 25], and are adapted for the use case of graphs. The convolution-types are based on studies with point clouds [26–28]. The physics features are concatenated at the global pooling layer. This is finally fed into a multi-layer perceptron (MLP) with ln(A) as the target output.

3.2 Primary Type Discrimination using GNNs

For a good mass discrimination, it is required that in all the energy bins the network-predictions for any two elements are maximally separated from each other. As can be seen from Figure 4, the GNN based method shows good mass resolution and preciseness in prediction over almost all the energy bins (true MC energy shown here). p-Fe separation shows improvement with increase in energy. The overlap-area between normalized p-Fe KDEs (Figure 4) is a around 27% for $6.4 \leq \log_{10}(\text{E/GeV}) < 6.6$ and improves to around 15% in the last-bin. For intermediate primaries there is still scope for improvement.

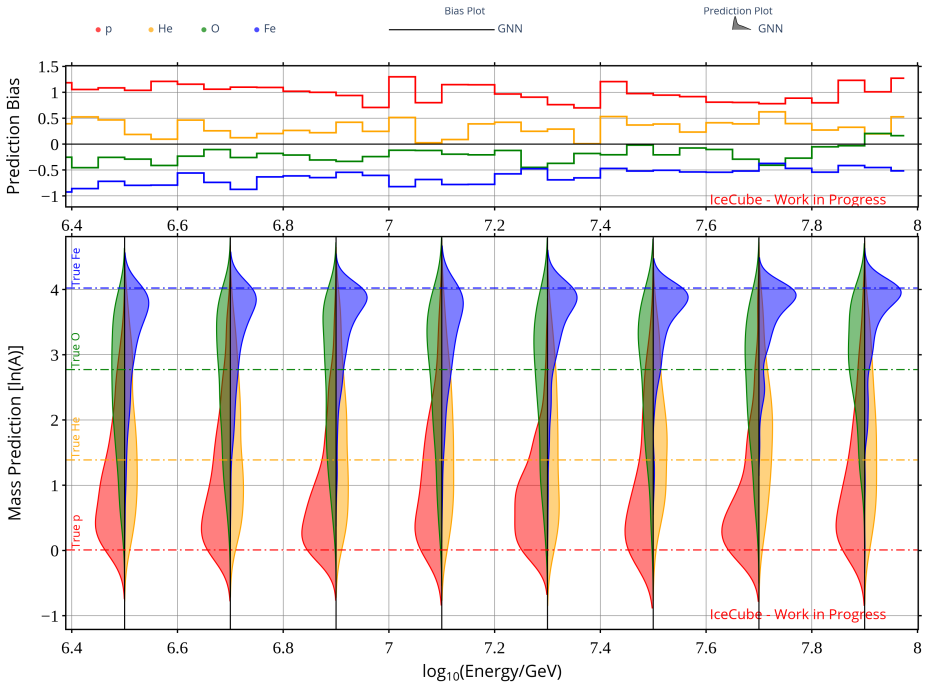

Figure 4: Kernel density estimate (KDE) [Bottom] of mass prediction by GNN in true MC energy bins and Horizontal dashed-dot lines represent True Primary Mass [ln(A)]. Top plot shows bias in predictions.

4. Conclusion and Outlook

We presented a graph neural network based approach for estimating the mass of primary cosmic rays in the energy range $5.0 \leq \log_{10}(\mathrm{E/GeV}) \leq 8.0$. The new results are promising and allow to report individual spectra for elemental groups in the future. There are also ongoing efforts to increase the discrimination-power of the newer composition sensitive parameters. Since graphs can be used for irregular-shaped data this study can in the future also be adapted for the planned multi-component irregular-shaped IceCube-Gen2 and the related surface-enhacement [17].

References

[1] M.G. Aartsen et al., *The IceCube Neutrino Observatory: instrumentation and online systems*, *Journal of Instrumentation* **12** (2017) P03012.

[2] R. Abbasi et al., *IceTop: The surface component of IceCube*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **700** (2013) 188.

[3] R. Abbasi, Y. Abdou, T. Abu-Zayyad, M. Ackermann, J. Adams, J. Aguilar et al., *The design and performance of IceCube DeepCore*, *Astroparticle physics* **35** (2012) 615.

[4] M. Aartsen et al., *Cosmic ray spectrum and composition from PeV to EeV using 3 years of data from IceTop and IceCube*, *Physical Review D* **100** (2019) 082002.

[5] P. Koundal, M. Plum and J. Saffer, IceCube collaboration, *Study of mass composition of cosmic rays with IceTop and IceCube*, *PoS* **ICRC2021** (2021) 323 [2107.09626].

[6] P. Koundal, IceCube collaboration, *Composition Analysis of cosmic-rays at IceCube Observatory, using Graph Neural Networks*, in *Workshop on Machine Learning for Cosmic-Ray Air Showers (2022), Newark, DE, USA, 31.01. 2022–03.02. 2022*, 2022.

[7] D. Heck et al., *CORSIKA: A Monte Carlo code to simulate extensive air showers*, *Report fzka* **6019** (1998) .

[8] G. Battistoni, F. Cerutti, A. Fasso, A. Ferrari, S. Muraro, J. Ranft et al., *The FLUKA code: Description and benchmarking*, in *AIP Conference proceedings*, vol. 896, pp. 31–49, American Institute of Physics, 2007.

[9] E.-J. Ahn, R. Engel, T.K. Gaisser, P. Lipari and T. Stanev, *Cosmic ray interaction event generator SIBYLL 2.1*, *Physical Review D* **80** (2009) 094003.

[10] S. Klepser, *Reconstruction of extensive air showers and measurement of the cosmic ray energy spectrum in the range of 1-80 PeV at the South Pole*, .

[11] H. Pandya, *Search for PeV gamma rays and astrophysical neutrinos with IceTop and IceCube*, University of Delaware (2019).

[12] A. Medina, *Reconstruction of Xmax and Energy from 3–100 PeV using 5 Years of Data From IceTop and IceCube and its Applications*, Ph.D. thesis, The Ohio State University, 2021.

[13] T.K. Gaisser, *Spectrum of cosmic-ray nucleons, kaon production, and the atmospheric muon charge ratio*, *Astroparticle Physics* **35** (2012) 801.

[14] J.C. Arteaga Velázquez and C. Andrea, *Study of the muon size and the charged number ratio in air showers as a mass sensitive parameter in KASCADE-Grande*, in *International Cosmic Ray Conference*, vol. 1, p. 271, 2011.

[15] W. Apel et al., *KASCADE-Grande measurements of energy spectra for elemental groups of cosmic rays*, *Astroparticle Physics* **47** (2013) 54.

[16] M.G. Aartsen et al., *IceCube-Gen2: the window to the extreme Universe*, *Journal of Physics G: Nuclear and Particle Physics* **48** (2021) 060501.

[17] A. Haungs, *A scintillator and radio enhancement of the IceCube surface detector array*, in *EPJ Web of Conferences*, vol. 210, p. 06009, EDP Sciences, 2019.

[18] P. Koundal, IceCube collaboration, *Graph Neural Networks and Application for Cosmic-Ray Analysis*, *PoS* **DLCP2021** (2022) 004.

[19] R. Abbasi et al., *A convolutional neural network based cascade reconstruction for the IceCube Neutrino Observatory*, *Journal of Instrumentation* **16** (2021) P07041.

[20] J. Micallef, IceCube collaboration, *Using convolutional neural networks to reconstruct energy of GeV scale IceCube neutrinos*, *Journal of Instrumentation* **16** (2021) C09019.

[21] S. Yu, IceCube collaboration, *Direction reconstruction using a CNN for GeV-scale neutrinos in IceCube*, *Journal of Instrumentation* **16** (2021) C11001.

[22] IceCube collaboration, J. Saffer, "Low-Energy Cosmic Ray Composition with IceCube and IceTop." Vortrag gehalten auf DPG-Frühjahrstagung der Sektion Materie und Kosmos (SMuK 2022), Online, 21.–25. März 2022, 2022.

[23] J. Zhu et al., *Beyond homophily in graph neural networks: Current limitations and effective designs*, *Advances in Neural Information Processing Systems* **33** (2020) 7793.

[24] C. Szegedy et al., *Going deeper with convolutions*, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015.

[25] G. Huang et al., *Densely connected convolutional networks*, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, 2017.

[26] H. Deng, T. Birdal and S. Ilic, *Ppfnet: Global context aware local features for robust 3d point matching*, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 195–205, 2018.

[27] H. Zhao, L. Jiang, J. Jia, P.H. Torr and V. Koltun, *Point transformer*, in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 16259–16268, 2021.

[28] Y. Wang, Y. Sun, Z. Liu, S.E. Sarma, M.M. Bronstein and J.M. Solomon, *Dynamic graph cnn for learning on point clouds.(2018), arXiv preprint arXiv:1801.07829* **222** (2018) .