EPJ Web of Conferences **319**, 13012 (2025) https://doi.org/10.1051/epjconf/202531913012
RICAP-24

Transfer Learning in KM3NeT/ORCA for Neutrino Event Reconstruction

Iván Mozún-Mateo[1,*] on behalf of the KM3NeT Collaboration

[1]LPC CAEN, Normandie Univ, ENSICAEN, UNICAEN, CNRS/IN2P3, 6 boulevard Maréchal Juin, Caen, 14050 France

Abstract. This study explores using transformer models to analyze data from the KM3NeT/ORCA neutrino telescope. Due to the increasing detector's size, reconstructing neutrino events is challenging. By training models on simulations for the full detector (115 detection units) and fine-tuning them on smaller configurations, significant performance improvements can be achieved compared to models trained from scratch on limited data samples. This approach also helps estimate the detector's sensitivity as it grows.

1 Introduction

A transformer model is a novel deep learning architecture that handles sequential data, processing their components in parallel, able to capture complex patterns and relationships on it [1]. Transfer learning is the technique of adapting models pre-trained on a specific task to solve a different but related problem [2]. In this study, we explore the potential of such methods to improve the reconstruction performances of the KM3NeT/ORCA neutrino telescope [4], currently in construction in the Mediterranean sea, at a depth of 2450 km about 40 km from Toulon (France), which aims to measure the neutrino mass hierarchy using atmospheric neutrinos in the range of 1-100 GeV [3]. The detector is a tridimensional array of photomultiplier tubes within digital optical modules arranged along vertical detection units (DUs). The light pattern can be arranged as a time-ordered sequence with position, direction, and timing information from the telescope within a time window of [-250, 2500] ns around the reconstructed neutrino interaction time, and can be used to reconstruct neutrino physics.

2 Transformers in KM3NeT/ORCA: first results and perspectives

Simulations for already deployed configurations of the telescope are generated following a run-by-run approach, i.e. using time-dependent data-taking conditions to generate MC simulated runs that account for the detector behavior in a given configuration. However, those runs are not always enough to properly train the models. When applying transfer learning from the full detector (115 DUs) to the ORCA6 and ORCA10 configurations (with 6 and 10 DUs respectively), we observe a significant improvement in event topology classification with limited data (1k training events) and limited training time ($\mathcal{O}$(1s) per epoch) with respect to models trained from scratch, as illustrated in Fig. 1 by the AUROC

---

[*] e-mail: mozun@lpccaen.in2p3.fr

value, i.e. the model's ability to distinguish the two topologies (track/shower), with higher values indicating better discrimination and robustness.

Even if the training sample increases, models trained from scratch are never able to overcome or to even reach comparable performances as those achieved by models fine-tuned on the full configuration, which retain valuable information from DUs not yet deployed. As an example, Fig. 2 shows the improvement in the reconstructed neutrino direction (left) and the performance boost (right) obtained for the two detector configurations, with a sample of only 200k training events.

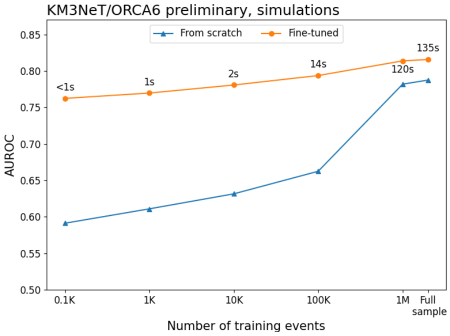

Figure 1: Evolution of the AUROC value for track/shower classification as a function of the training data size in ORCA6 for a model fine-tuned from ORCA115 (orange line) and a model trained from scratch (blue line).

The opening angle plateaus at around 5° for 6 DUs. At this stage, neutrino reconstruction performances are limited both by the small size of the detector (<10% of its final configuration) and by the training sample. In the near future, this method will be applied to a reconstruction algorithm trained on all ORCA configurations that have been running so far (up to ORCA24), to further improve the characterization of neutrino events in KM3NeT/ORCA.

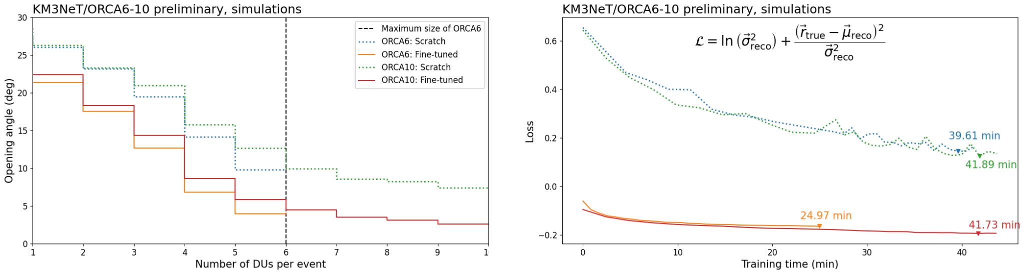

Figure 2: Left: opening angle in neutrino direction reconstruction (as estimated by the number of DUs containing at least one optical module having collected data) for ORCA6 and ORCA10 configurations, comparing models trained from scratch and models fine-tune from ORCA115 detector. Right: Gaussian negative log likelihood loss function time-evolution for the various models.

References

[1] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin, Attention Is All You Need. CoRR **abs/1706.03762**, (2017), http://arxiv.org/abs/1706.03762

[2] Jacob Devlin, Ming-Wei Chang, Kenton Lee-Kristina Toutanova, BERT, Pre-training of Deep Bidirectional Transformers for Language Understanding. CoRR **abs/1810.04805**, (2018), http://arxiv.org/abs/1810.04805

[3] KM3NeT collaboration, Determining the neutrino mass ordering and oscillation parameters with KM3NeT/ORCA. The European Physical Journal C **abs/2103.09885**, (2022), https://arxiv.org/abs/2103.09885

[4] KM3NeT collaboration, Letter of Intent for KM3NeT 2.0. Journal of Physics G: Nuclear and Particle Physics **abs/1601.07459**, (2022), https://arxiv.org/abs/1601.07459