Enhancing low energy reconstruction and classification in KM3NeT/ORCA with transformers

Iván Mozún Mateo[1*] on behalf of the KM3NeT collaboration

1 Laboratoire de Physique Corpusculaire de Caen

⋆ mozun@lpccaen.in2p3.fr

The 2nd European AI for Fundamental Physics Conference (EuCAIFCon2025) Cagliari, Sardinia, 16-20 June 2025

Abstract

The current KM3NeT/ORCA neutrino telescope, still under construction, has not yet reached its full potential in neutrino reconstruction capability. When training any deep learning model, no explicit information about the physics or the detector is provided, thus they remain unknown to the model. This study leverages the strengths of transformers by incorporating attention masks inspired by the physics and detector design, making the model understand both the telescope design and the neutrino physics measured on it. The study also shows the efficacy of transformers on retaining valuable information between detectors when doing fine-tuning from one configurations to another.

1 Introduction

KM3NeT/ORCA is a neutrino telescope [1] placed in the Mediterranean sea, at a depth of 2450 km about 40 km from Toulon (France), with the objective of measuring the neutrino mass hierarchy using atmospheric neutrinos [2]. The detector design is a tridimensional array of PhotoMultiplier Tubes (PMTs) hosted within Digital Optical Modules (DOMs) [3] arranged along vertical detection units (DUs). The telescope collects light from Cherenkov photons emitted along the path of charged particles in neutrino interactions, creating a time-ordered sequence of light pulses with position and timing information that can be used to reconstruct neutrino event kinematics.

In this study, a novel deep learning architecture named transformer [4], that handles sequential data as those observed in a neutrino telescope is presented. In section 2, the model is introduced and the use of attention mask inspired by physics and detector design is motivated. In section 3, the challenges of reconstructing neutrino physics, both due to the physics itself and the limited telescope size, and how transformers overcome them are described.

2   A transformer model for neutrino telescopes

Thanks to the light pattern originated from a neutrino interaction, transformers are very well suited for reconstructing physics in neutrino telescopes. A transformer model handles sequential data, processing their components in parallel, being able to capture complex patterns and relationships on it through the use of the self-attention mechanism [4],

$$\texttt{Attention}(Q, K, V) = \texttt{SoftMax}\left(\frac{Q \cdot K^T}{\sqrt{d}}\right) \cdot V \tag{1}$$

where Q, K and V are matrices built from the input data, and d is the latent space size of the model. If no prior information is given to the model, eq. (1) provides information about how the components of the input sequence are correlated among themselves. Nevertheless, no information about the physics is directly injected in the model.

To overcome this burden, physics constraints and detector information are artificially included in eq. (1) through the use of attention masks U [5,6],

$$\texttt{PairwiseAttention}(Q, K, V; U) = \texttt{SoftMax}\left(\frac{Q \cdot K^T}{\sqrt{d}} + U\right) \cdot V \tag{2}$$

These attention masks are $N \times N$ matrices with N being the input sequence length, that encode different correlations between the light pulses of an event, for instance, the space-time relativistic distance to identify pulses from the same source [6], the euclidean distance to quantify spatial proximity, or a local-coincidence mask to determine which pulses come from the same PMT, DOM or DU.

3   Challenges of neutrino physics reconstruction in KM3NeT/ORCA

When talking about reconstruction in a neutrino telescope, the main difficulty found is based on its detection principle. Since the PMTs detects light, only charged particles can easily be reconstructed, whereas non-charged particles are invisible to it, leading to an intrinsic bias when trying to reconstruct the whole interaction.

3.1   A growing telescope

The KM3NeT/ORCA telescope grows in size by adding DUs to already deployed ones, increasing its capability and sensitivity to measure neutrino properties. When a deep learning model is initialized, it does not contain physics information, it learns as it trains. However, this is not optimal because a model can already have learnt that information from a previous configuration. Besides, if a model is first trained on a larger telescope it retains valuable information from DUs not yet deployed as shown in fig. 1 by the area under the Receiver Operating Characteristic (ROC) curve, a metric that indicates the model's ability to distinguish the two neutrino events: ν_μ^{CC}/ν_e^{CC}.

Figure 1 shows an improvement of over 20% achieved with a very limited training sample (100 events per class), by using interpolated information from a larger configuration, contrary to a model trained from scratch, whose performances are only comparable when using a large training sample (1M events per class), where limitations from the detector itself play a higher role rather than the model capabilities.
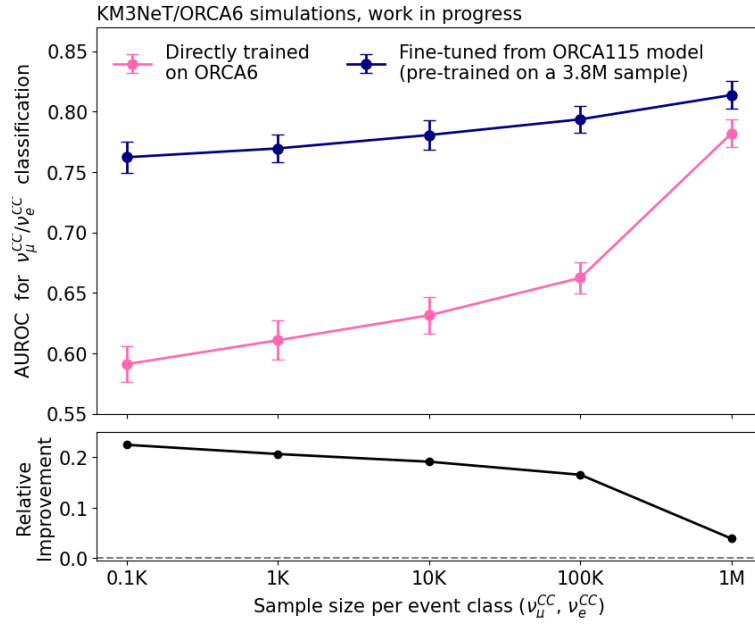
Figure 1: Evolution of the AUROC value for ν_μ^{CC}/ν_e^{CC} classification as a function of the training data size in ORCA6 for a model fine-tuned from ORCA115 (purple line) and a model trained from scratch (pink line). Modified from [7].

Another benefit of the use of deep learning for classification is that the training does not rely on any reconstructed variable, it computes the score it from what the raw data. On the contrary, classical classification methods rely on decision-tree-based classifiers trained using reconstruction variables. These techniques are prone to be strongly biased by the reconstruction algorithm used and must be updated whenever a new version of the reconstruction software is released.

3.2 Reconstructing neutrino energy and direction from light

The major difficulty of neutrino telescopes is that they are built to measure neutrino properties, but neutrinos are invisible to them. Therefore, reconstruction algorithms based on a maximum-likelihood fit (MLF) are developed to reconstruct only what is observed, which in most cases is based on either a track or a shower hypothesis, whereas in reality it is a mix of many things for which a likelihood is hard to define. Another limitation also arises from the fact that non-visible particles are involved, leading to an intrinsic bias when reconstructing the neutrino energy from the outgoing particles.

A advantage of transformers compared to MLF algorithms is that they can be trained to fit any hypothesis, which means they can, in principle, reconstruct anything observed by the telescope. Also, if we train the transformer on all types of events[1], it can handle complex cases better. For instance, a high-energy ν_μ^{CC} event that has both tracks and stochastic energy losses in the form of showers. Since the transformer has seen both tracks and showers during training, it can learn to reconstruct them more accurately whenever they appear in a new event and the inference time is significantly reduced since one single reconstruction is considered.

Therefore, as long as the neutrino properties are present in the training data, a deep learning model is trained to reconstruct them and to establish the necessary likelihood to extrapolate the neutrino properties from the observations in the telescope. Figure 2 demonstrates the improvement in direction (over 20%) and energy estimation w.r.t. a MLF algorithm when using transformers.

---

[1]For this work, only CC interaction processes are considered.
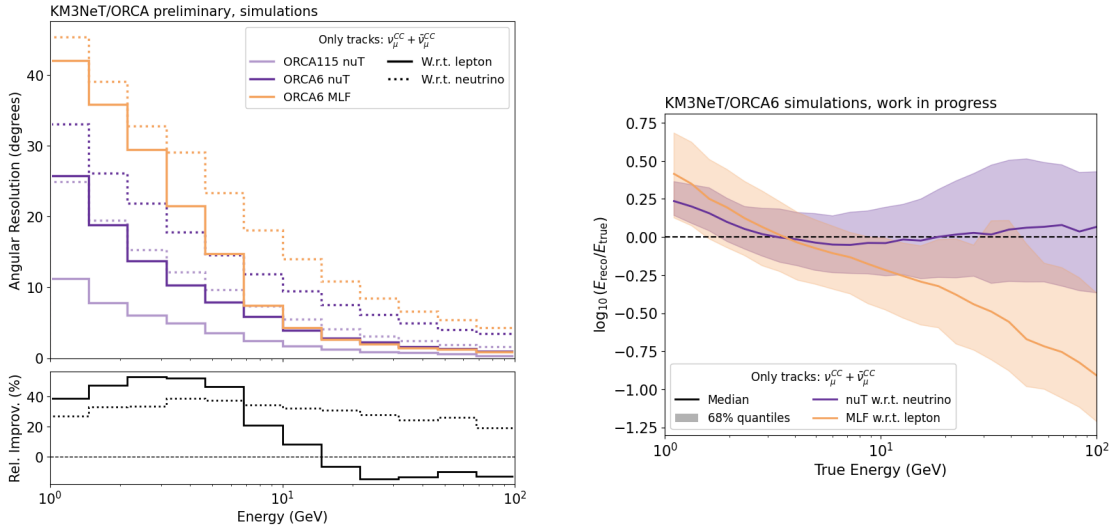
Figure 2: Left: angular resolution in direction reconstruction for ORCA6 (dark purple) and ORCA115 (light purple). Right: log 10 of ratio between reconstructed and true energy. Comparison to a MLF reconstruction algorithm is also shown in orange color.

4   Conclusion

Transformers and the use of, physics and detector-inspired attention masks, significantly improve neutrino reconstruction in KM3NeT/ORCA. Compared to maximum-likelihood fits algorithms, they achieve better direction and energy resolution of the neutrino at low energies. The use of pre-trained models in larger configurations retains essential physics information and reduces the cost of training models as the detector evolves.

References

[1] S. Adrián-Martínez et al. (The KM3NeT collaboration), *Letter of intent for KM3NeT 2.0*, *Journal of Physics G: Nuclear and Particle Physics* **43**, 084001 (2016), http://dx.doi.org/10.1088/0954-3899/43/8/084001.

[2] S. Aiello et al. (The KM3NeT collaboration), *Determining the neutrino mass ordering and oscillation parameters with KM3NeT/ORCA*, *The European Physical Journal C* **82**, 1 (2022), http://dx.doi.org/10.1140/epjc/s10052-021-09893-0.

[3] S. Aiello et al. (The KM3NeT collaboration), *The KM3NeT multi-PMT optical module*, *Journal of Instrumentation* **17**, P07038 (2022), doi:10.1088/1748-0221/17/07/P07038.

[4] A. Vaswani et al., *Attention Is All You Need*, *SciPost Phys.* (2023), https://arxiv.org/abs/1706.03762.

[5] H. Qu, C. Li, S. Qian, *Particle Transformer for Jet Tagging*, arXiv:2202.03772 (2024), https://arxiv.org/abs/2202.03772.

[6] H. Bukhari et al., *IceCube – Neutrinos in Deep Ice: The Top 3 Solutions from the Public Kaggle Competition*, arXiv:2310.15674 (2023), https://arxiv.org/abs/2310.15674.

[7] I. Mozún-Mateo, *Transfer Learning in KM3NeT/ORCA for Neutrino Event Reconstruction*, *EPJ Web Conf.* **319**, 13012 (2025), doi:10.1051/epjconf/202531913012.