“Deep Learning in KM3NeT”

Simona Maria Stellacci (INFN) for the KM3NeT Collaboration

*Cristiano Bozza, Rosa Coniglione, Chiara De Sio, Thomas Eberl, Steffen Hallmann,
Michael Moser, Stefan Reck*

“ Reconstruction and Machine Learning in Neutrino Experiments”
16-18 September, Hamburg

KM3NeT Infrastructure

KM3NeT is a neutrino research infrastructure located in the Mediterranean sea.

<u>Main features:</u>

• Wide energy range: 3GeV→10PeV

• Full sky coverage with the best sensitivity for galactic source

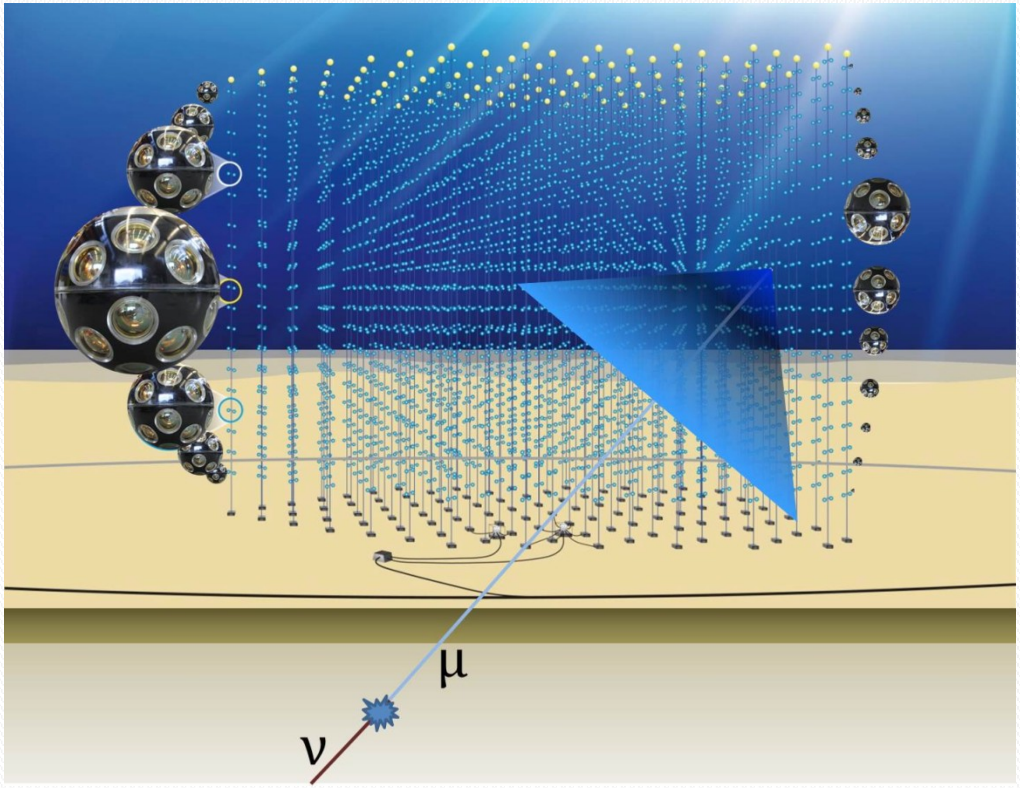
• All-flavuor neutrino detection

• Good angular resolution (<0.5° for muons; ~ 2° for electrons at high energy)

KM3NeT Experiment

Main Goals:

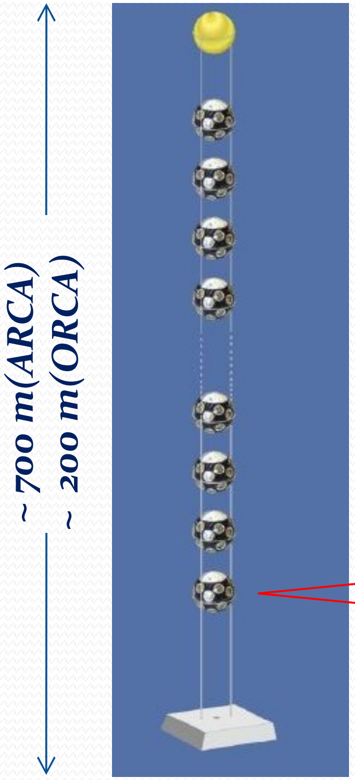
- *identification of high energy astrophysical neutrino sources*
- *determination of neutrino mass hierarchy*
- *supernova core collapse*
- *Dark Matter*

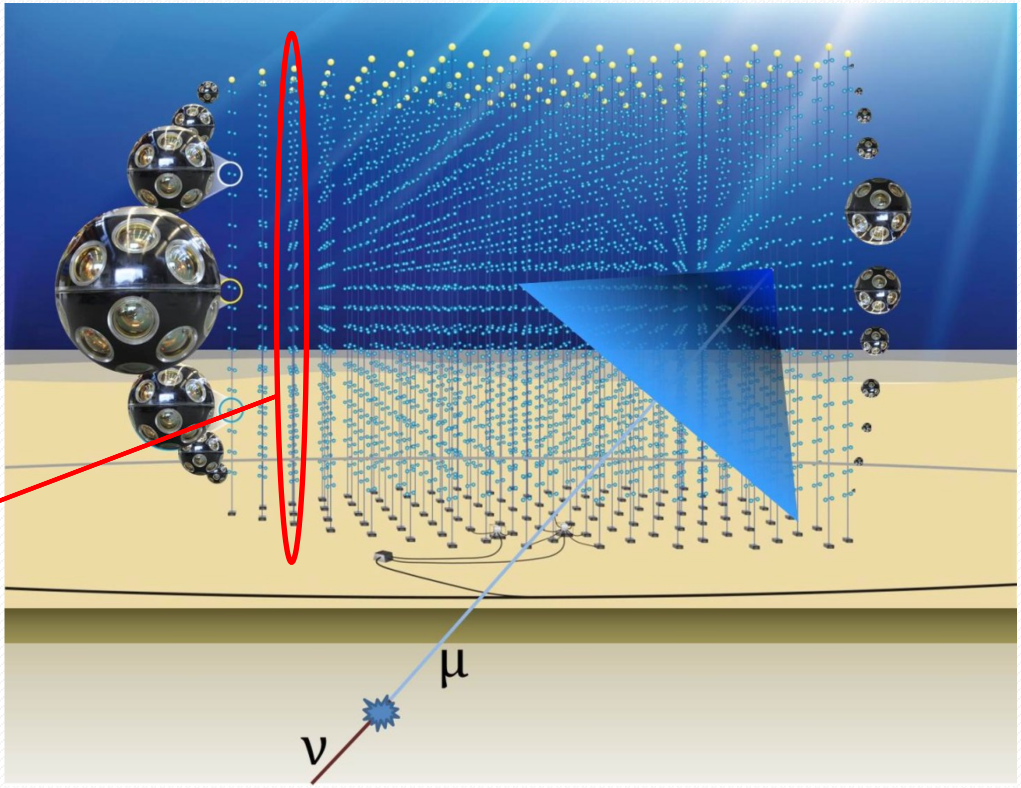


Detection principle : *the neutrinos are detected by measuring the Cherenkov light emitted by secondary particle generated in neutrino interaction*

KM3NeT Detector

Detection Unit*: vertical string hosting 18 DOMs anchored at seabed and taut by buoys.*





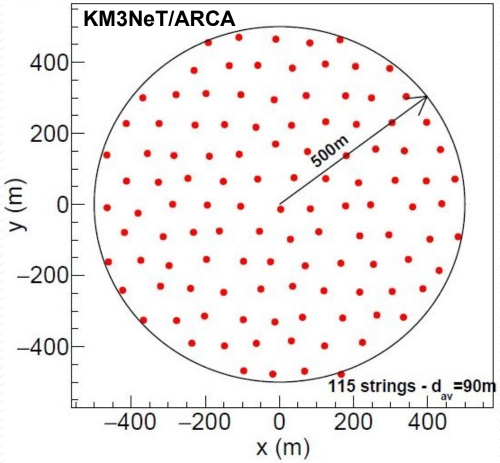
DOM*: 17" glass sphere hosting 31 PMTs*

KM3NeT Site

KM3NeT/ARCA

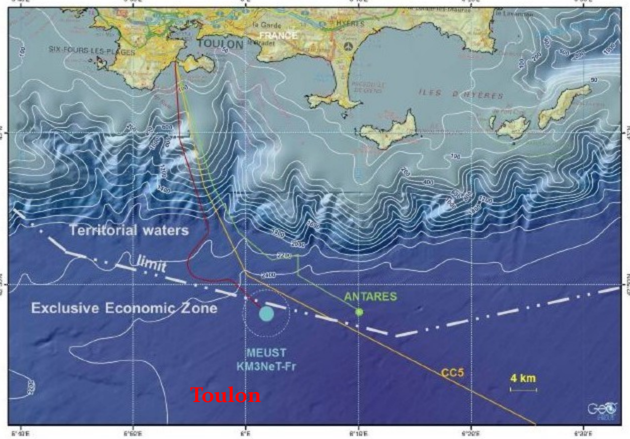
Astroparticle Research with Cosmic in the Abyss

- *2 "building blocks" of 115 Detection Units each*
- *700 m in height*
- *Depth 3500m*
- *DOMs spaced 90m in X-Y, 36 m in Z*

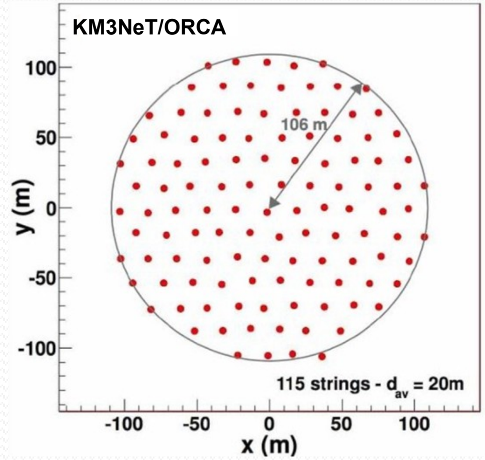


KM3NeT/ORCA

Oscillation Research with Cosmic in the Abyss



- *1 "building blocks" of 115 Detection Units*
- *200 m in height*
- *Depth 2500m*
- *DOMs spaced 20m in X-Y, 9 m in Z*



Deep Learning applications in KM3NeT/ARCA

Deep Learning applications in KM3NeT/ARCA

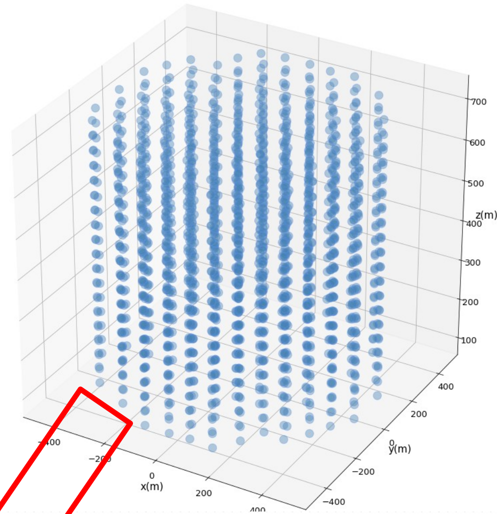
Four Convolutional Neural Network (CNN) models designed for :

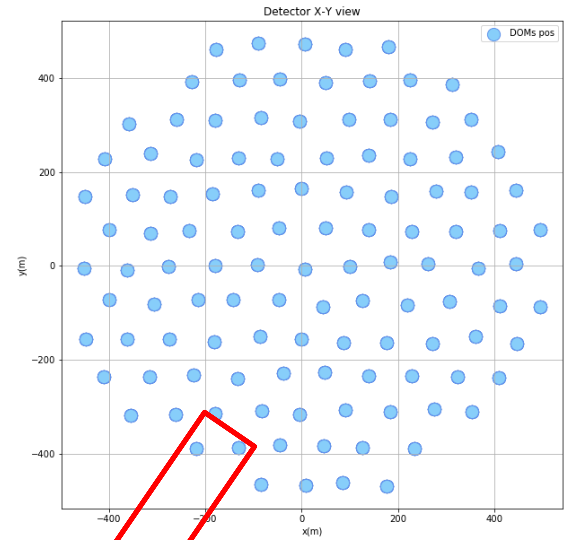
- ***Classification***
  - *up-going / down – going incident particle*
  - ν_μ *CC /* ν_e *CC interaction*
- ***Regression***
  - *particle energy estimation*
  - *particle direction estimation*

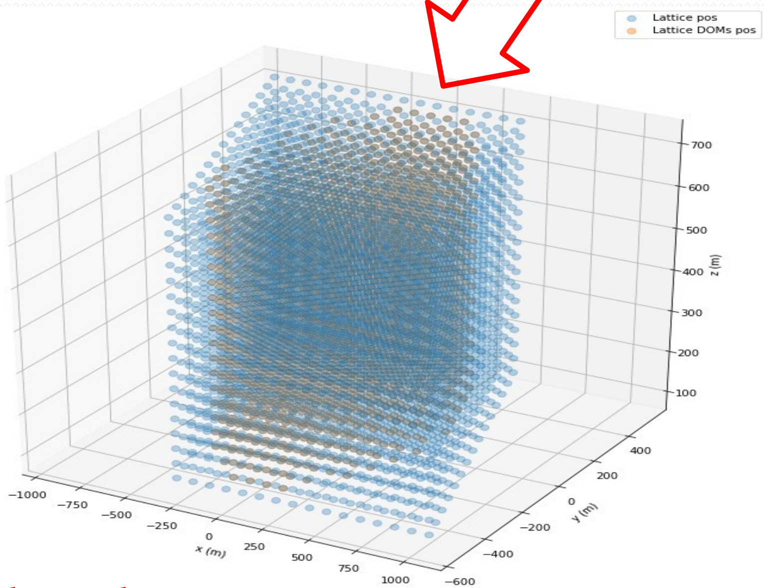
Data Preparation

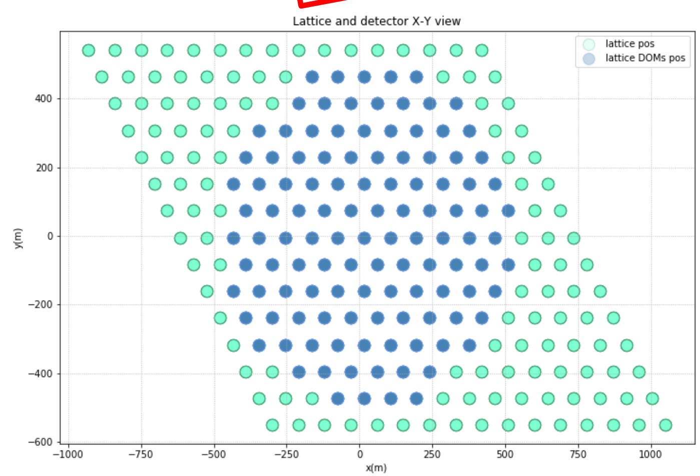
<u>*Regularised Detector Structure*</u>

- *exactly 90m spaced in (X,Y)*
- *exactly 36m spaced in Z*
- *regularised detector mapped in* ***lattice***









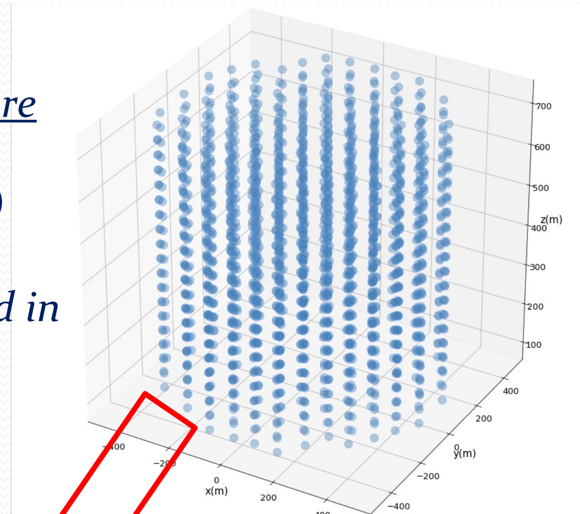
Regularized Detector XYZ-View

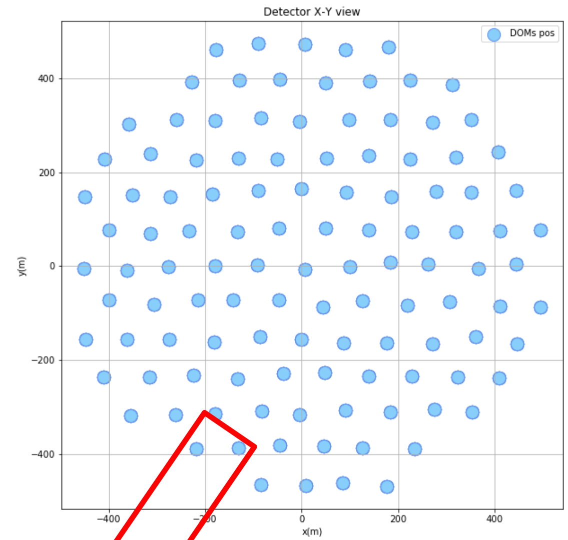
Regularized Detector XY-View

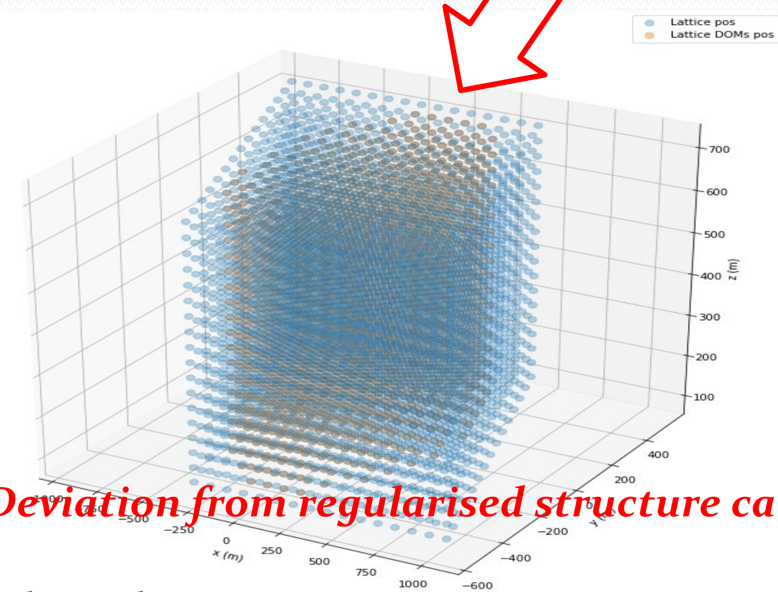
Data Preparation

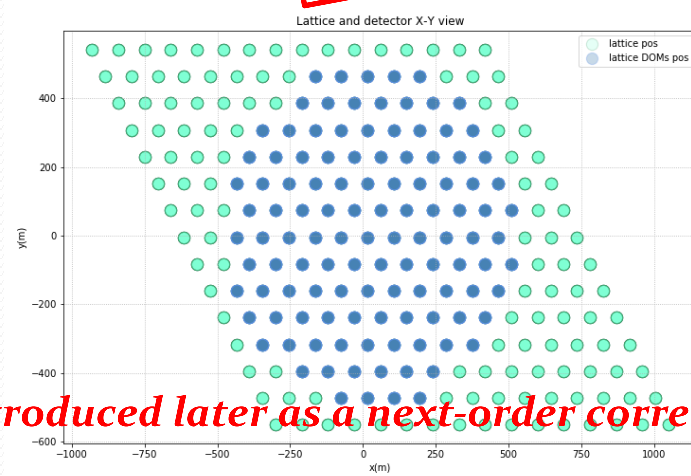
<u>*Regularised Detector Structure*</u>

- *exactly 90m spaced in (X,Y)*
- *exactly 36m spaced in Z*
- *regularised detector mapped in* ***lattice***









Deviation from regularised structure can be introduced later as a next-order correction

Regularized Detector XYZ-View

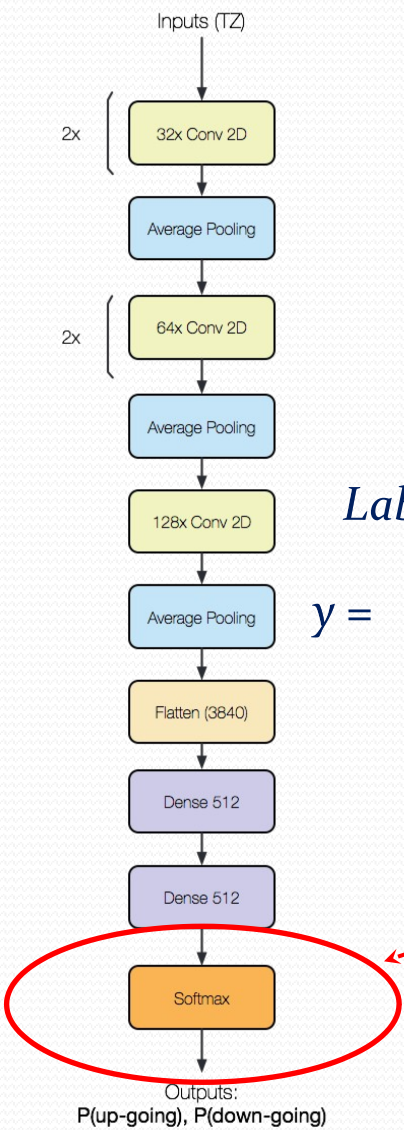
Regularized Detector XY-View

DATASET

258,879 total events

NVIDIA GEFORCE GTX-1080Ti GPU

Up/Down – going Classification



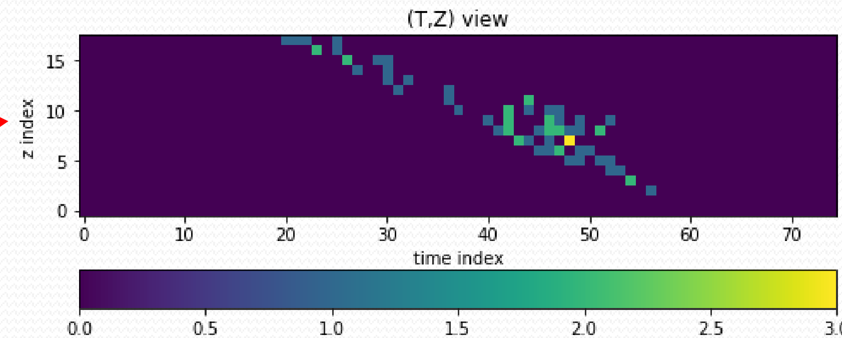
__Model Architecture:__

- *3 x 2D Convolutional block*
  - *2 x 2D Conv layers+ Average Pooling layers*
- *3 x Fully – Connected layers*

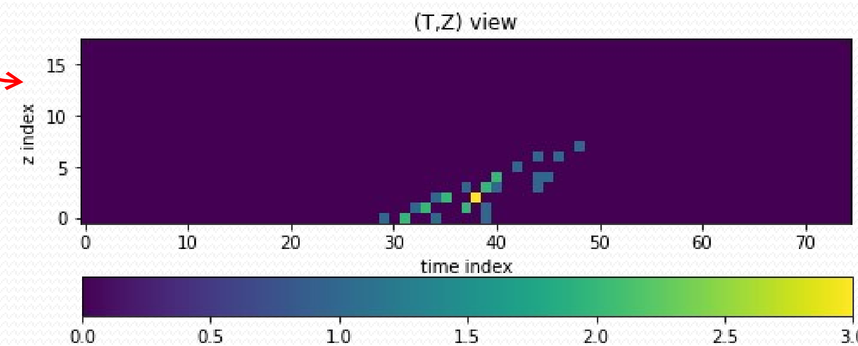
Labels

$$y = \begin{cases} \cos\theta \leq 0 \\ \cos\theta > 0 \end{cases}$$

Classification



Downgoing ($\cos\theta \leq 0$)

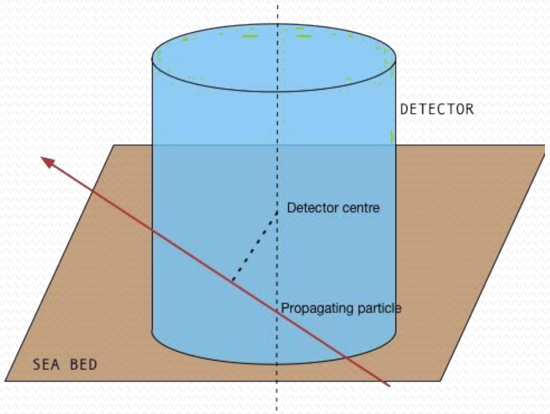


Upgoing ($\cos\theta > 0$)

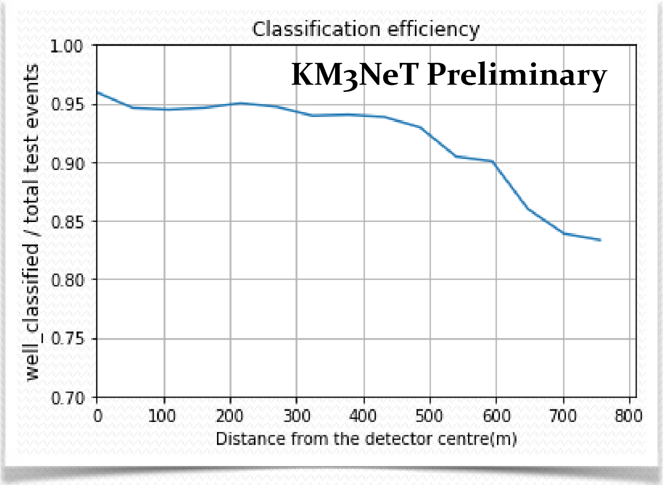
Up/Down – going Classification

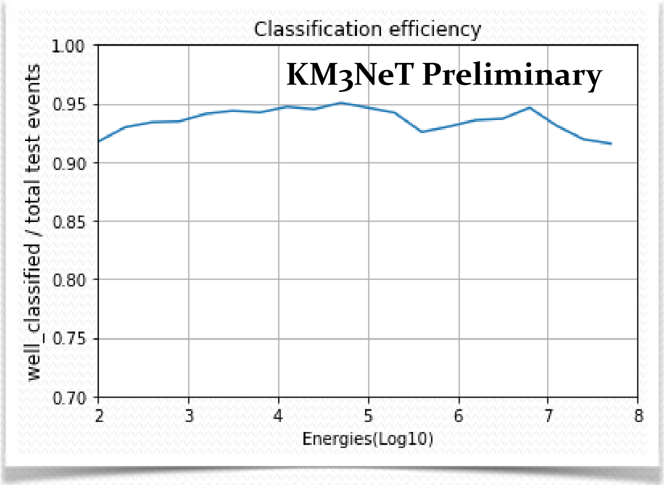
Performance depends on the track length in detector

Best performance :
- *high energy events*
- *short distances from centre of detector*

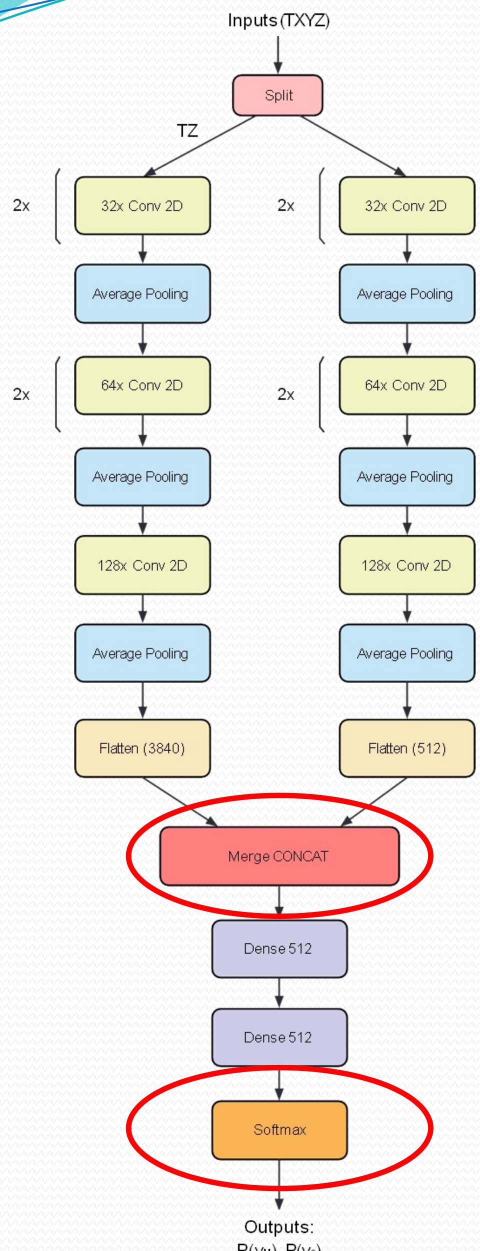


***Efficiency** : numbers of well - classified events/total test events*





Classification ν_μ CC - ν_e CC interactions

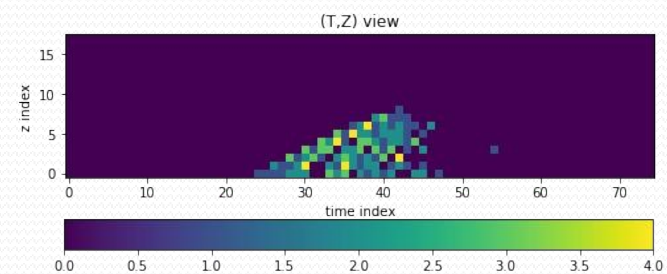


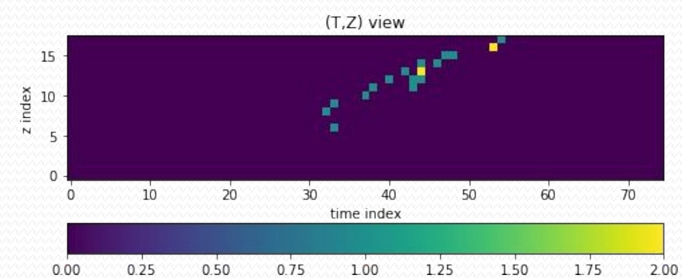
Model Architecture:

- *two separated branches*
  - *(Z,t) evolution*
  - *(X,Y) evolution*
- *3 x 2D Convolutional blocks*
- *merged to extract the common features*

Labels

$$y = \begin{cases} 1 & \nu_e CC \\ 0 & \nu_\mu CC \end{cases}$$

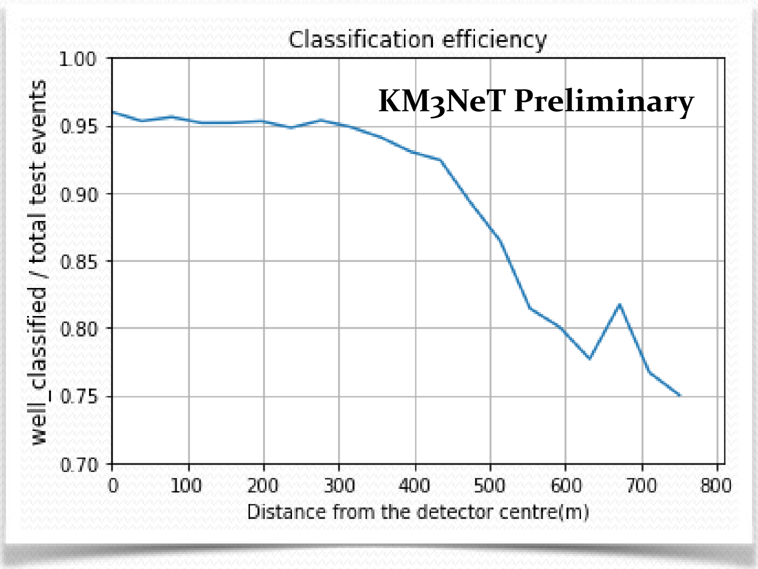


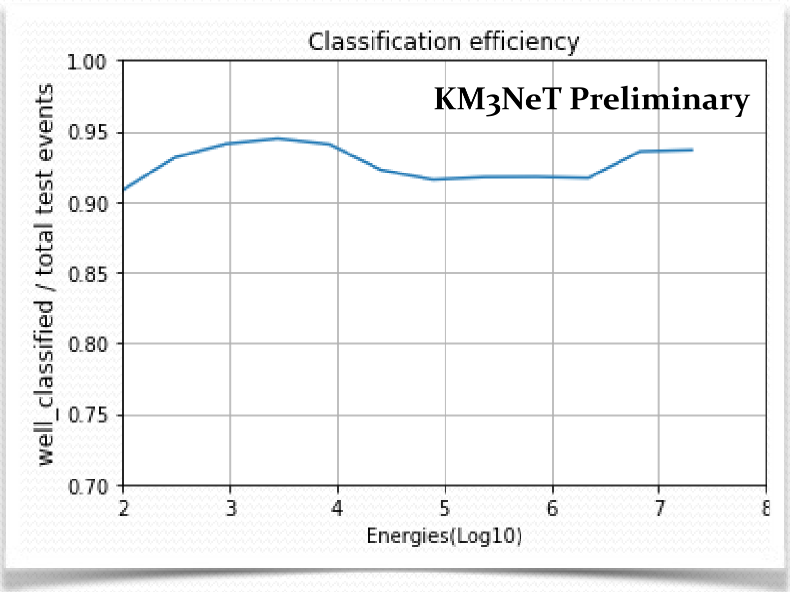


Classification ν_{μ} CC - ν_{e} CC interactions

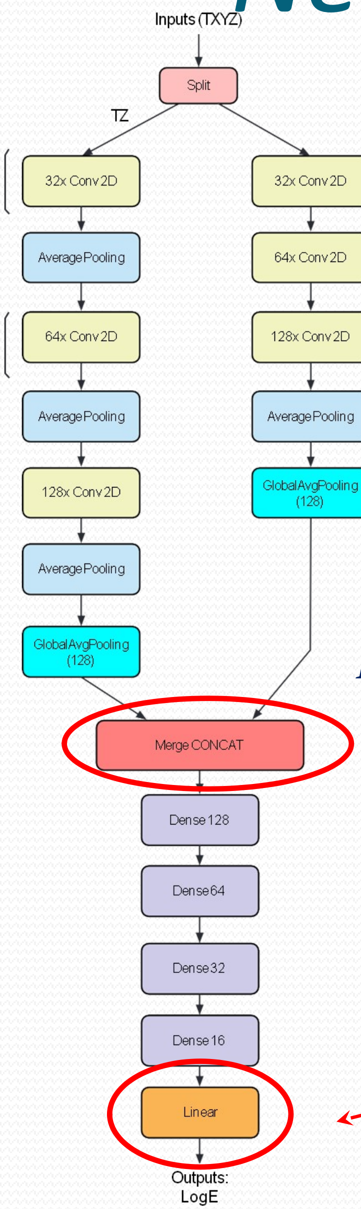
Best performance :

- *high energy events*
- *short distances from centre of detector*





Neutrino Energy Estimation



Model Architecture:

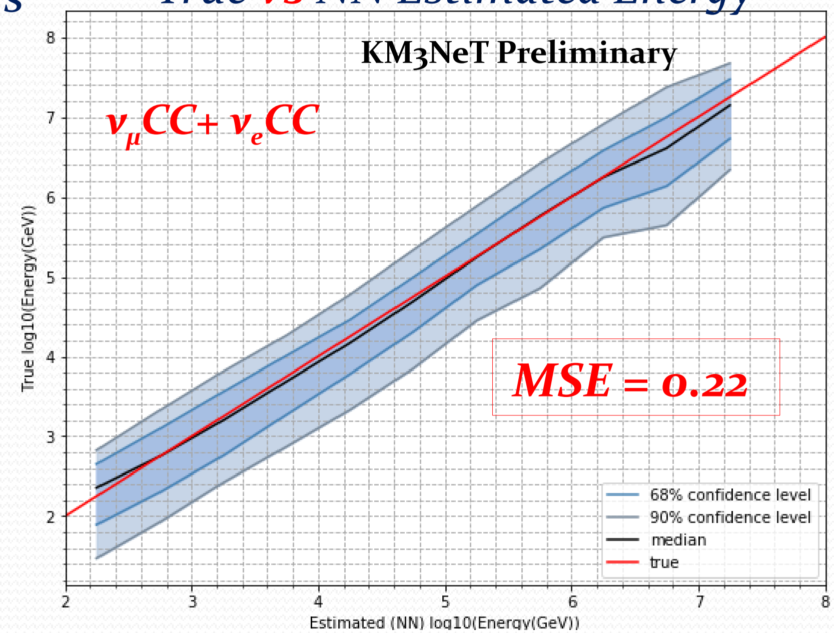
- two separate branches
  - (Z,t) evolution
  - (X,Y) evolution
- 2 x 2D Convolutional blocks concatenated
- 4 x fully – connected layers

Labels

$Log_{10}(Energy)$: MCtruth

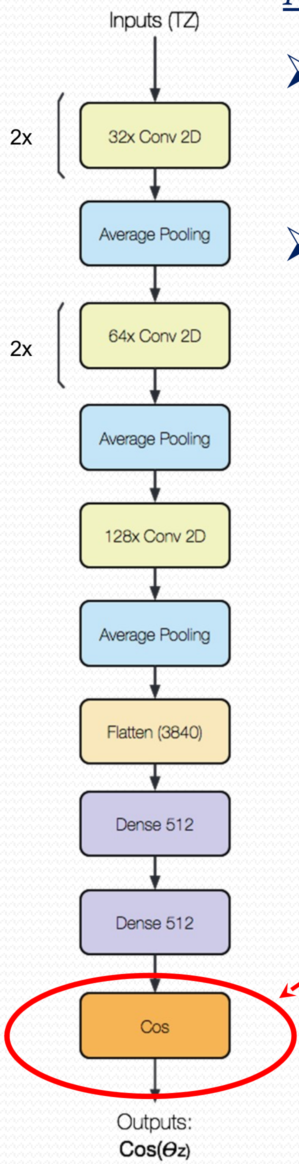
energy estimation

True **vs** NN Estimated Energy



- Good linearity
- Could be improved with more statistics at high energy

Neutrino Direction Estimation



Model Architecture:

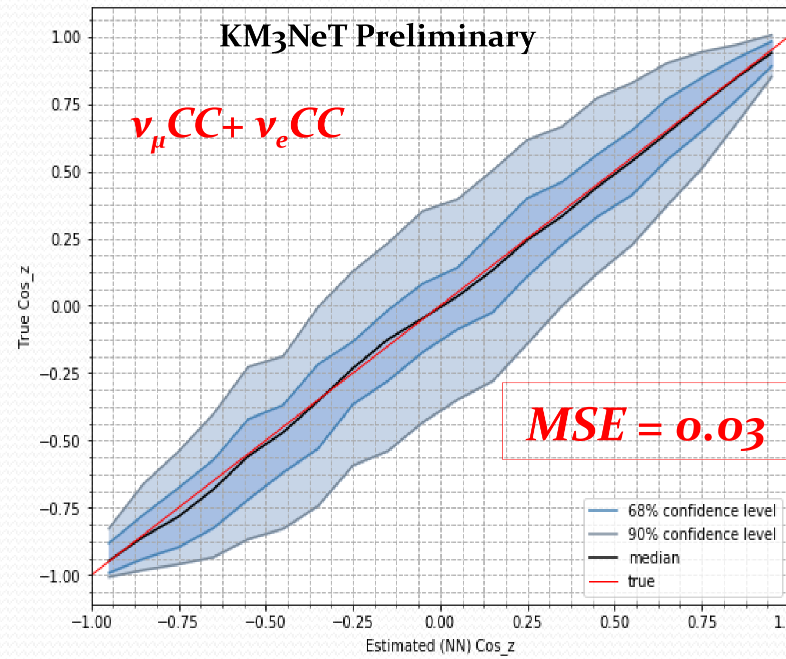
- *3 x 2D Convolutional block*
  - *(Z,t) evolution*
- *3 x Fully Connected layers*

Labels

$\cos\theta_z$: MCtruth

Activation Function

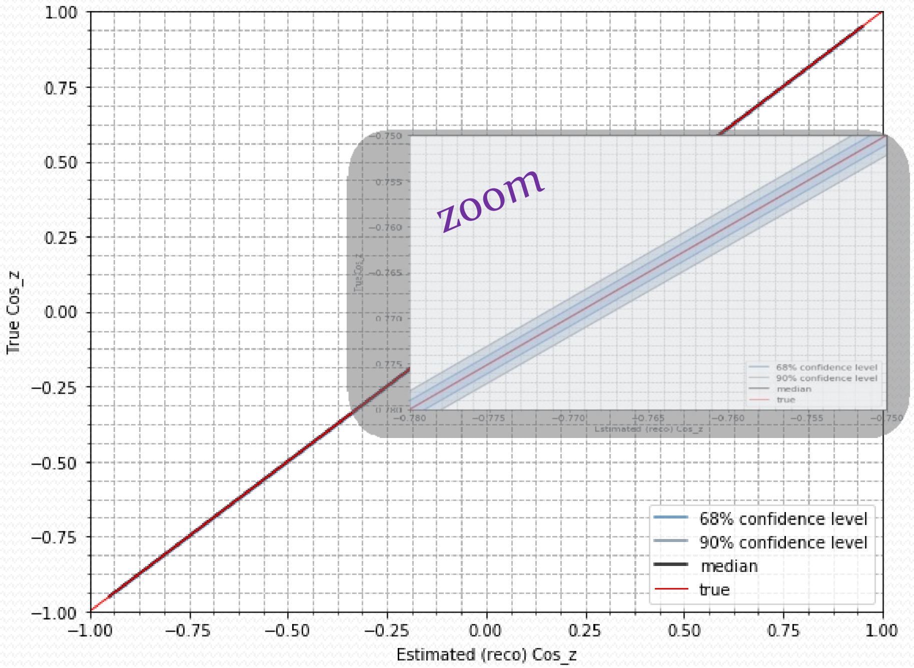
*True **vs** NN Estimated Direction(Z)*



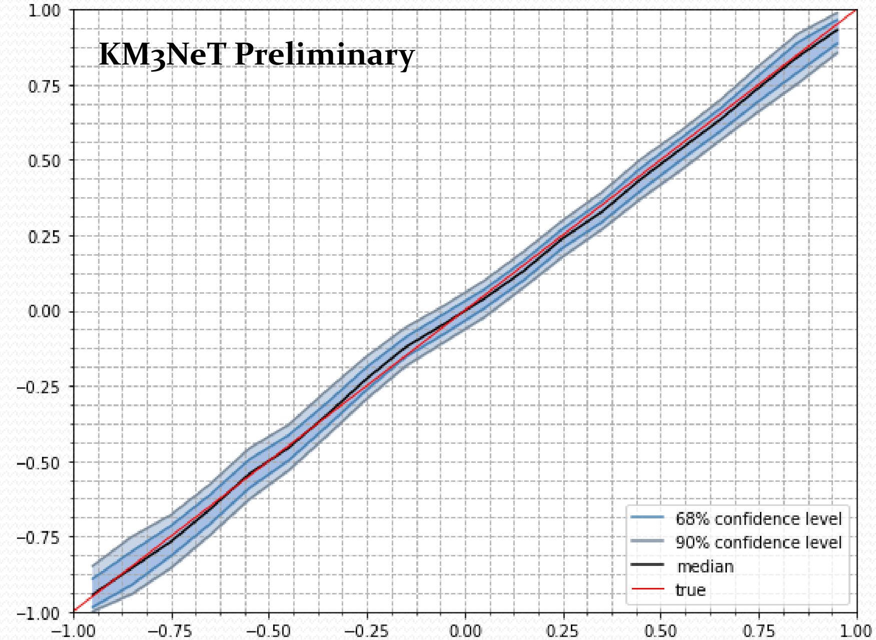
Comparison with conventional algorithms (KM3NeT/ARCA)

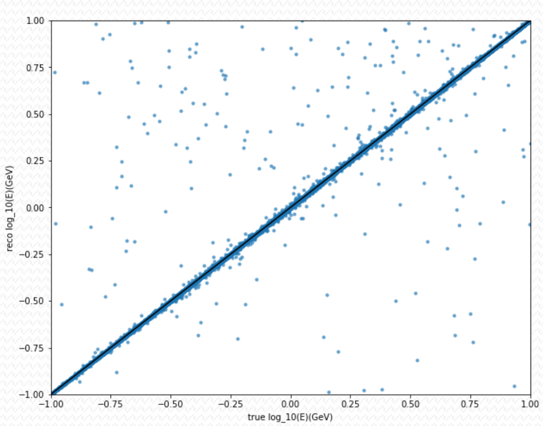
Direction Estimation

*True **vs** Std. Reconstruction Direction(Z)*

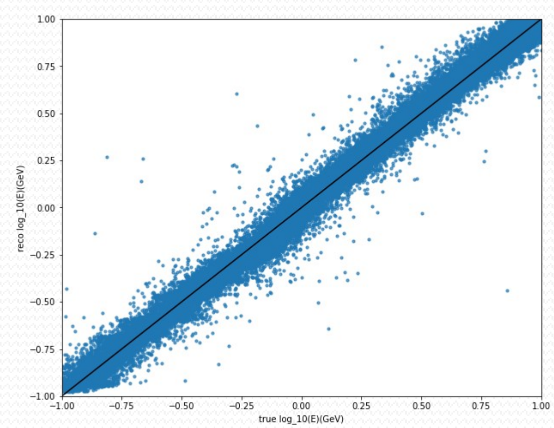


*True **vs** CNN Reconstruction Direction(Z)*





*Estimated **vs** True*
Neutrino direction



Comparison on Up/Down – Going Classification

- *apply "labels" to reconstructed events*
  - *$\cos(\theta_z) > 0$ "up-going"*
  - *$\cos(\theta_z) \leq 0$ "down-going"*
- *compare predictions*

Accuracy on Up/Down – Going Classification

KM3NeT standard reconstruction (-lik>60 and log(β_o)>-2.8)	*CNN classification running test on **ν_μCC** events*
Classification Accuracy	**Classification Accuracy**
0.998	**0.987**

*Only **ν_μCC** reconstructed events with quality cuts*

CNN always yields a direction, even when the conventional fit does not meet the quality requirements.

Deep Learning applications in KM3NeT/ORCA

Deep Learning applications in KM3NeT/ORCA

Convolutional Neural Network (CNN) models designed for :

- ***Classification***
  - *background suppression*
  - *track/shower – like event topologies*
- ***Regression***
  - *particle energy estimation*
  - *particle direction estimation*
  - *vertex position*

+

Error Estimate on Regression results

Model Architectures

Input

CONV 3D (3x3x3)

Batch Normalization

Activation ReLu

Max Pooling

Dropout

Output

end of network differs depending of the tasks

(X,Y,Z,P) → (X,Y,Z) grid + PMT identifier

+

(X,Y,Z,T) → (X,Y,Z) grid + time of hit recording

NVIDIA TESLA V100 GPU

\+ \+ Keras

Classification *: "softmax" classifier with 2 additional neurons to derivate classified output*

Regression *: 7 additional neurons with linear activation to regress continuous variables*

Background/Signal Classification

MonteCarlo sample:

- *neutrinos*
- *atmospheric muons*
- *random noise*

Data Set :

- *Training Set (75%)*
- *Validation Set (2.5%)*
- *Test Set (22.5%)*

CNN model architecture:
3D Conv blocks + 2 fully connected layers

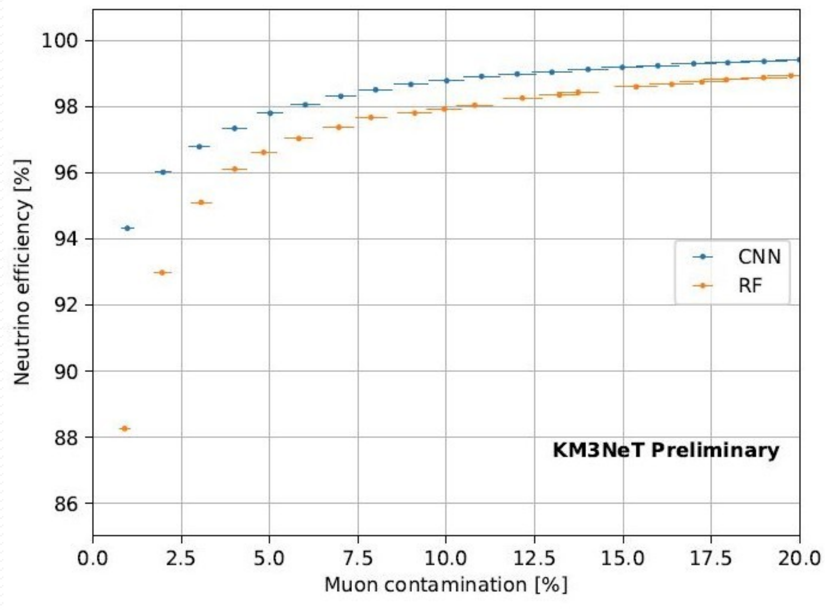
Output layer:
distinction between neutrino and no neutrino - events

Background/Signal Classification

RESULTS

The comparison is made using ONLY the events from the ***test set***

CNN background classifier ***vs*** *RF classifier*



discrimination threshold

$$C_\mu(p) = \frac{N_\mu(p)}{N_{tot}(p)}$$

atmospheric muon contamination

neutrino efficiency

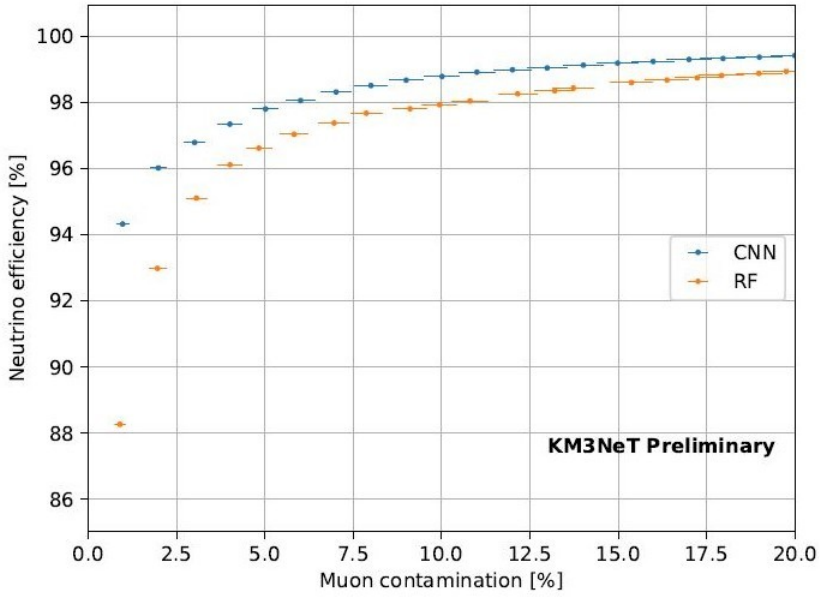
$$\nu_{eff}(p) = \frac{N_\nu(p)}{N_{\nu,tot}}$$

Background/Signal Classification

RESULTS

*The comparison is made using ONLY the events from the **test set***

*CNN background classifier **vs** RF classifier*



discrimination threshold

neutrino efficiency

$$C_\mu(p) = \frac{N_\mu(p)}{N_{tot}(p)} \qquad \nu_{eff}(p) = \frac{N_\nu(p)}{N_{\nu,tot}}$$

atmospheric muon contamination

*Performance of CNN classifier are **better** than those of RF classifier .*

*Neutrino efficiency is significantly **higher** when atmospheric muons are suppressed*

Track/Shower –like event Classification

Events are mostly neutrino events

Background events discarded by the CNN classifier

Track/Shower –like event Classification

Events are mostly neutrino events

Background events discarded by the CNN classifier

Sample:

- *50% track-like events ($\nu_\mu CC$)*
- *50% shower-like events*
  - *50% $\nu_e CC$*
  - *50% $\nu_e NC$*

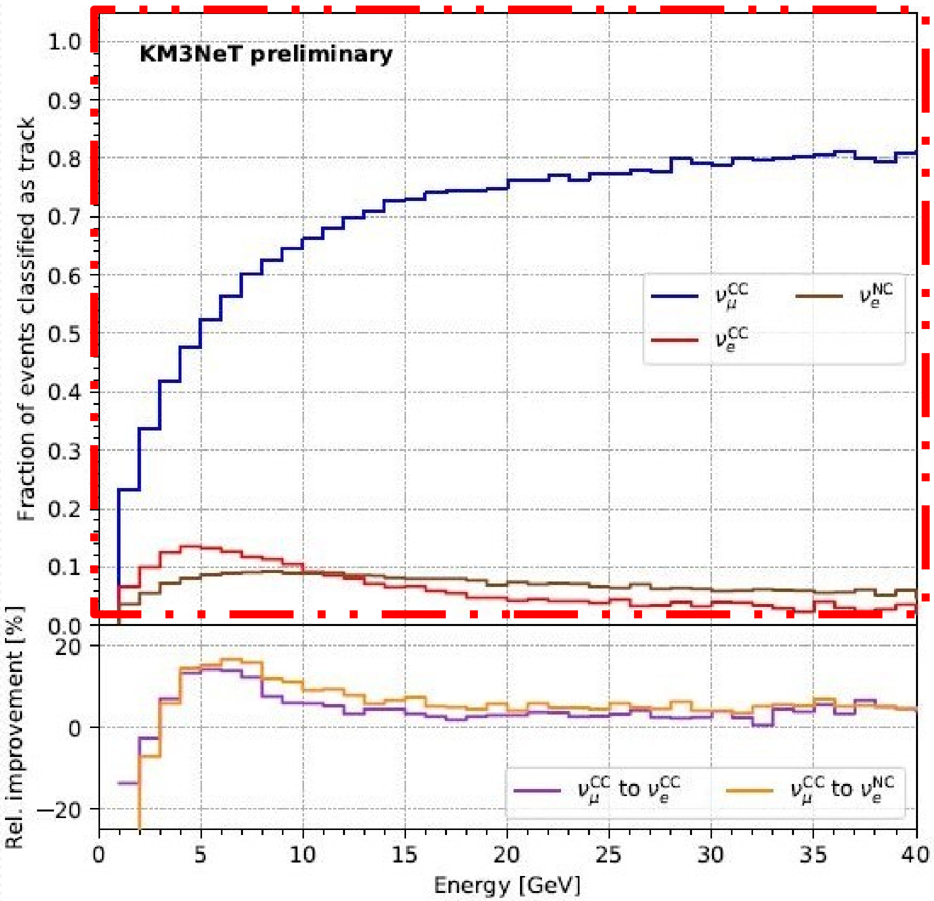
The dataset is "reweighted" to make the track/shower ratio indipendent of neutrino energy

Reweighted data set :

- *Training Set (70%)*
- *Validation Set (6%)*
- *Test Set (24%)*

Track/Shower –like event Classification

RESULTS

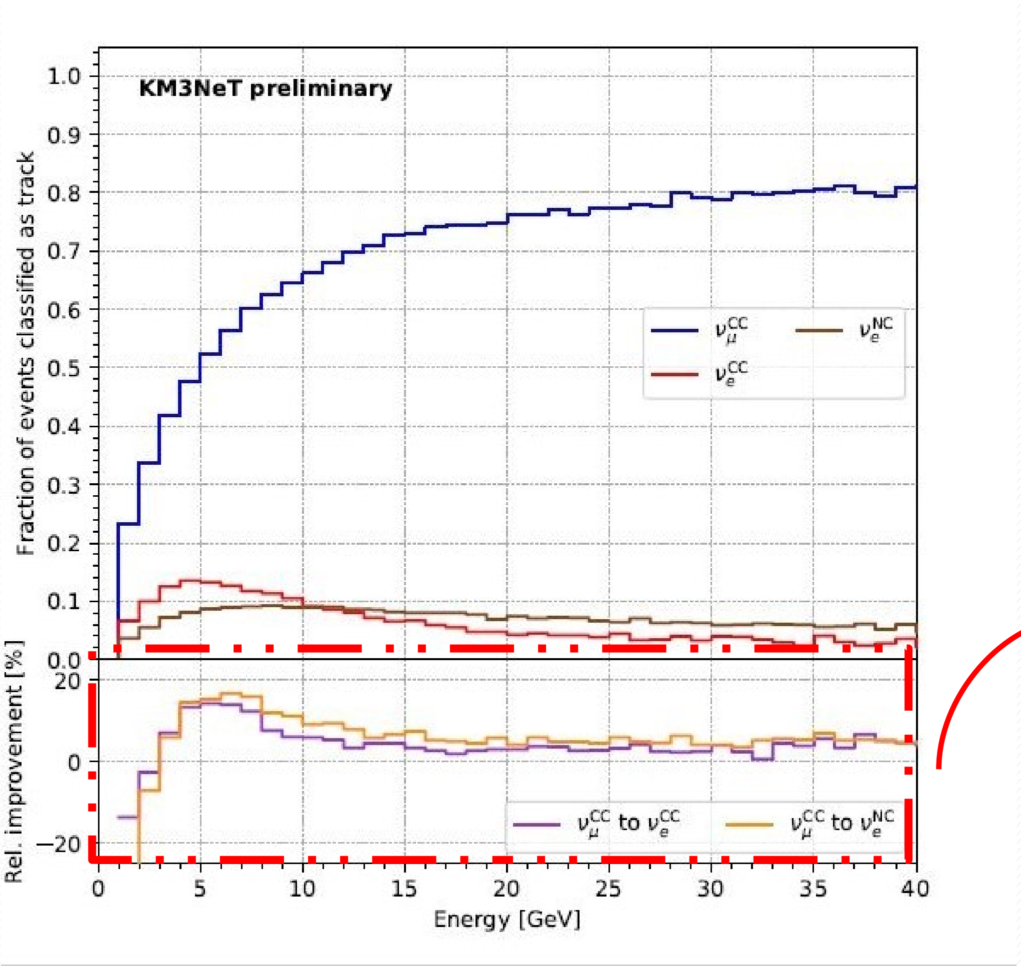


Event is classified as "track-like" events only if probability is ***≥ 0.5***

- Fraction of the events classified as "track-like" events
- 1 GeV < E_ν < 40 GeV
- Different interaction channels ($\nu_\mu CC$, $\nu_e CC$, $\nu_e NC$)

Track/Shower –like event Classification

RESULTS



- *The performance of the CNN classifier are better than the RF classifier.*
- *The performance gain is more significant for* $E_\nu < 15$ GeV

Event Parameter Regression

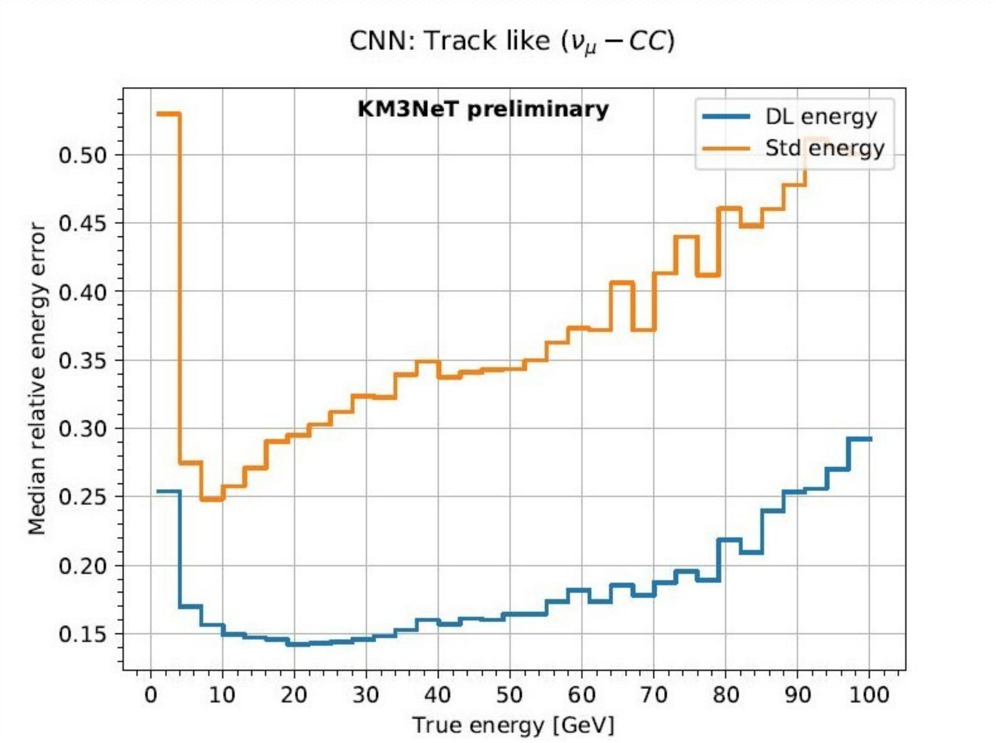
CNN network architecture:

3D Conv blocks + 1 fully connected layer

Properties Estimated :

- neutrino energy
- direction
- vertex position

Energy Fit

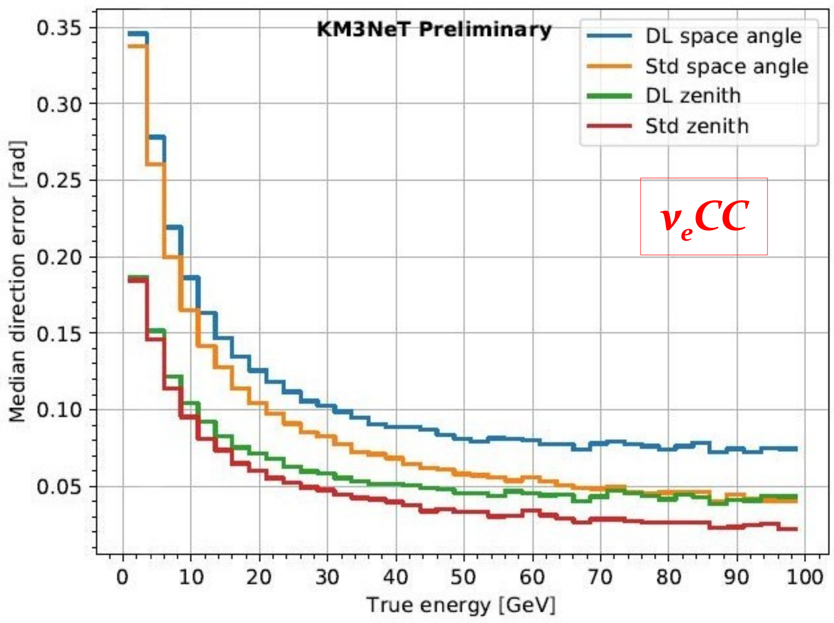


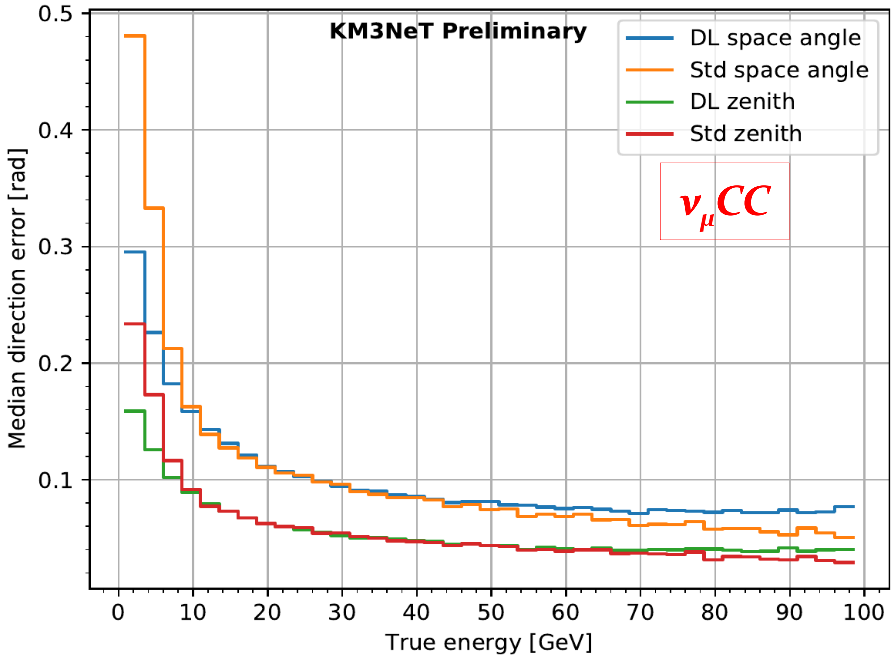
$$MRE = \langle \left| \frac{E_{reco} - E_{true}}{E_{true}} \right| \rangle$$

The performances of the CNN regressor with respect the Maximum Likelihood shower reconstructed algorithm are significantly better.

Direction Fit

The median absolute error (ME) is defined as the median of the distribution of residuals distribution for the reconstructed direction





*CNN **vs** Max Likelihood method:*

- *full direction*
- *zenith angle*

The performance of CNN are significantly better than the MaxLikelihood method for energies below a few GeV.

The performance of MaxLikelihood method improves with the energy.

Error Estimate

Output of the network → $\vec{y}_{reco}$

$\vec{y}_{reco}$ = *(E+dir*$_{reco}$*+vtx*$_{reco}$*))*

one additional neuron to estimate the uncertainty for each component of $\vec{y}_{reco}$

Loss Function

$$L = \frac{1}{n}\sum_{i=0}^{n}(\sigma_{reco} - |y_{true} - y_{reco}|^2)$$

reconstructed uncertainty

The network learns to estimate the average absolute residual $\sigma_{reco} \approx \langle y_{abs} \rangle$

Error Estimate

Output of the network → $\vec{y}_{reco}$

$\vec{y}_{reco} = (E + dir_{reco} + vtx_{reco}))$

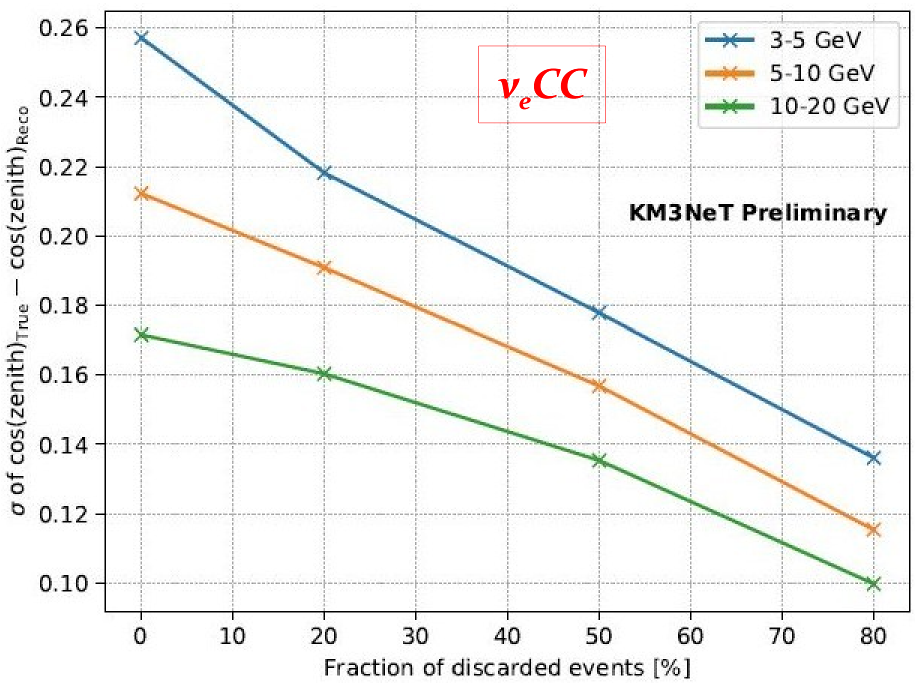
one additional neuron to estimate the uncertainty for each component of $\vec{y}_{reco}$

Loss Function

$$L = \frac{1}{n}\sum_{i=0}^{n}(\sigma_{reco} - |y_{true} - y_{reco}|^2)$$

reconstructed uncertainty

The network learns to estimate the average absolute residual $\sigma_{reco} \approx \langle y_{abs} \rangle$



The zenith angle resolution improves significantly when the events with the largest reconstruction error are discarded.

Conclusions and Outlook

- *Deep Learning models produce very promising results*

- *They provide stable estimations comparable with the conventional reconstruction algorithms*

Next Steps

- *Detailed detector description*

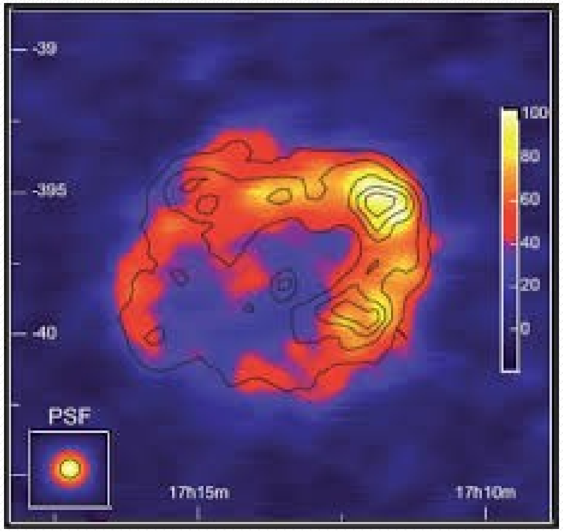
- *Add ν_τCC in topology classification*

- *Identification of neutrino flavour*

Back-up slides

KM3NeT/ARCA

Astroparticle Research with Cosmic in the Abyss



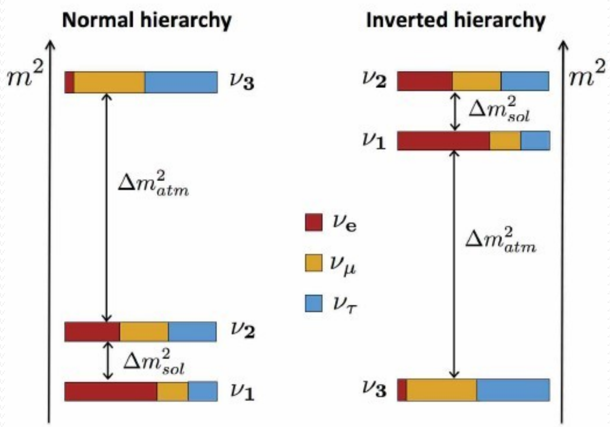
Search for point – like Galactic sources

High Energy neutrinos (100 TeV - 10 PeV):

- *Track –like events with angular resolution less than 0.2°*
- *Cascade – like events thanks to the good angular resolution (1° - 2°)*

KM3NeT/ORCA

Oscillation Research with Cosmic in the Abyss



- *Determination of Neutrino Mass Hierarchy*
- *Indirect Dark Matter search*
- *Track –like and Cascade-like events induced by "low" energy neutrinos (few GeV – 10 TeV)*

Comparison Settings

(ARCA)

- run "official" software (JGandalf) on the same test dataset used for CNN
- selected only ν_μCC events
- compare results on direction estimation
- compare performance on up/down – going selection