# ML HK Strategy

Thorsten Lux

Some more HK ML Thoughts

- Presented last week my doubts in respect of using hit time

- Also the "brute force" approach of giving a 2D image and the CNN should tell us everything seemed to be too optimistic

- Situation seems even worse since I wrongly assumed that CNN knows for each pixel/PMT position in xyz

- Time information without spatial information makes not so much sense => in 3D sum of red arrows in not equal to yellow arrow

- Each PMT provides:
  - Charge: q
  - Hit time: t
  - Position: x,y,z
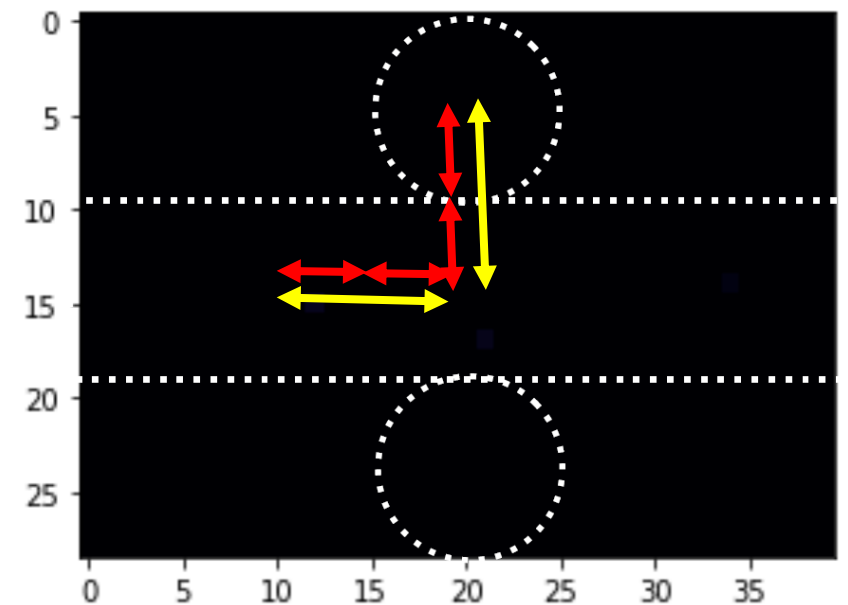
DUNE

- Focussing on 2D CNN might come from DUNE

- But LAr TPCs produce naturally 2D images and get 3D dimension from time

- DUNE much more advanced in ML reco

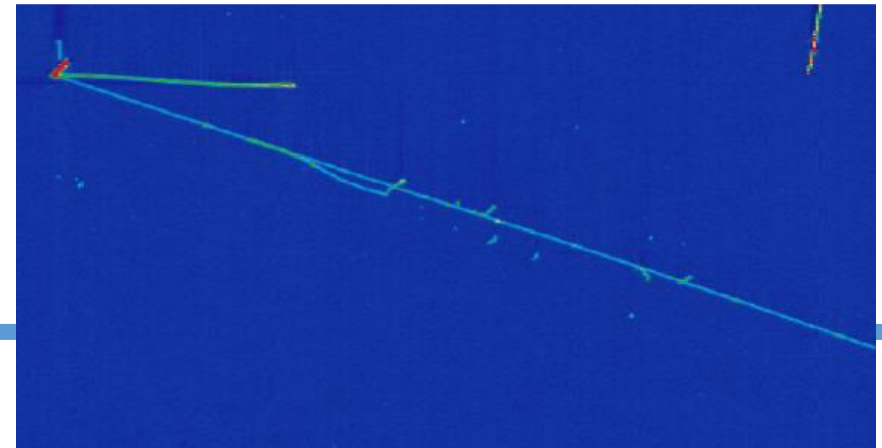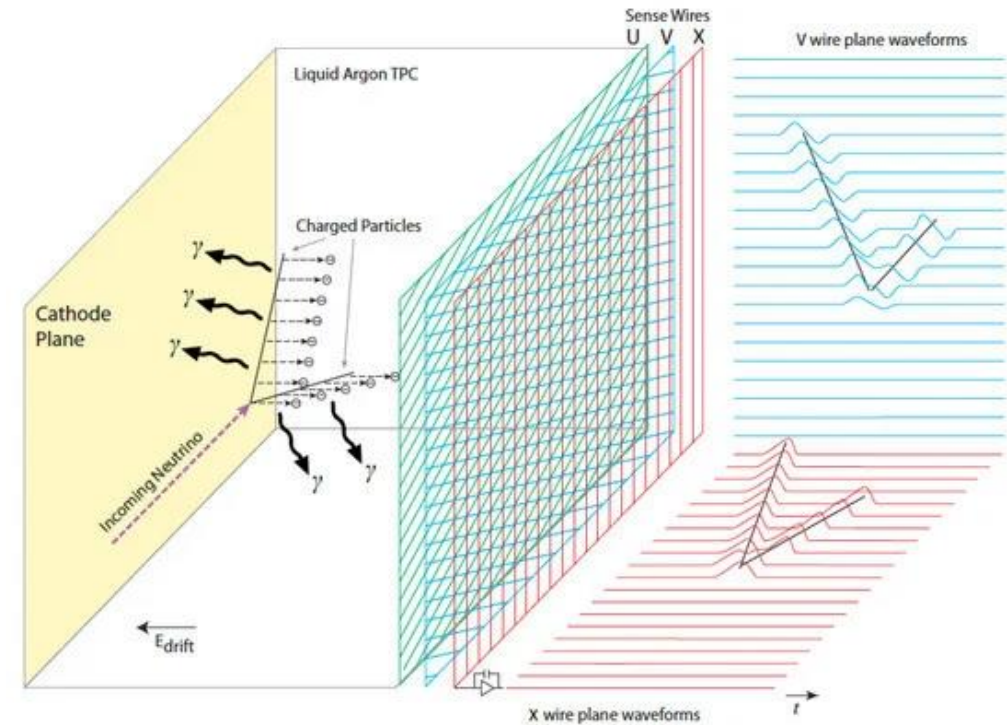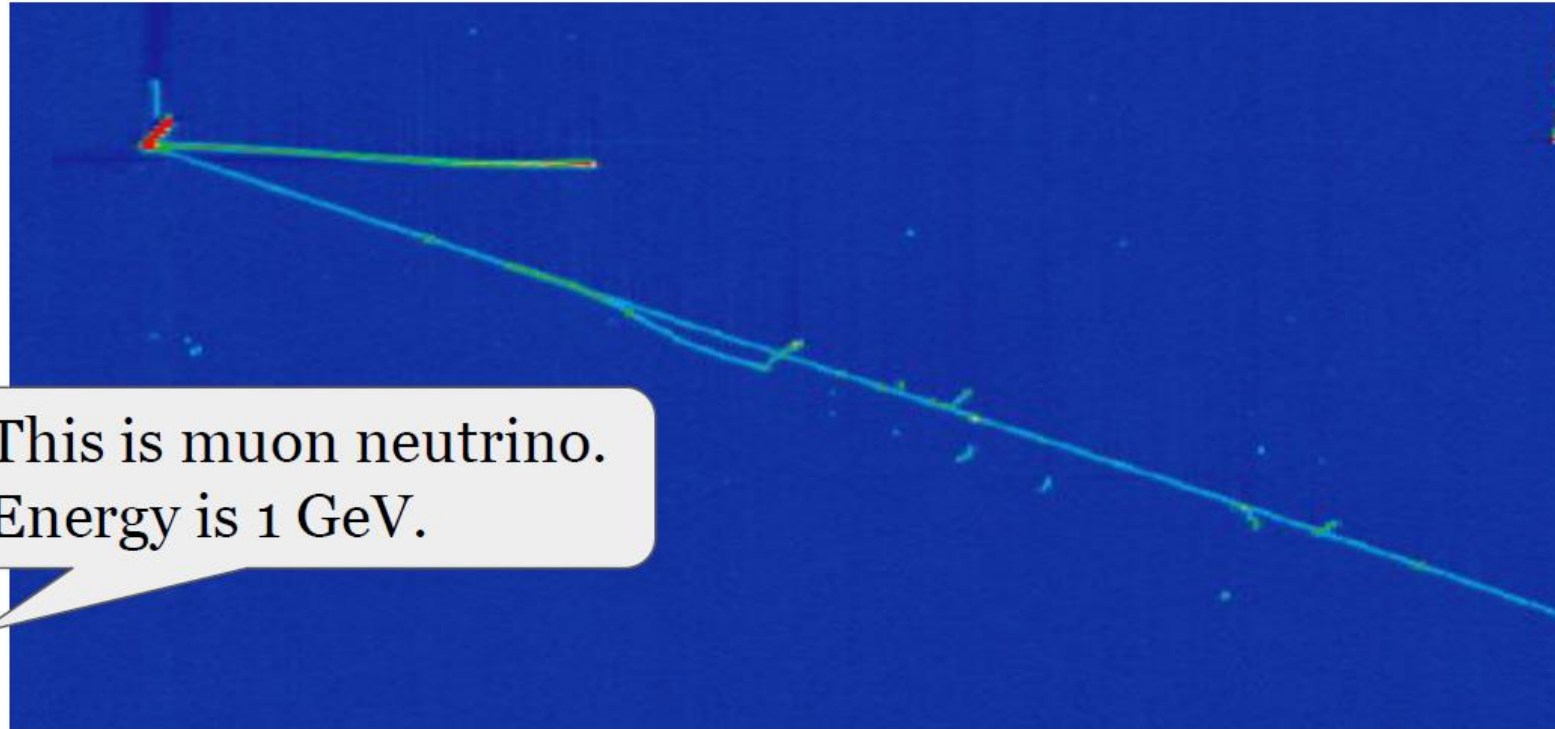- Situation for HK much more complex also from the geometry

Image classification/regression: straight to "flavour & energy"

This is muon neutrino. Energy is 1 GeV.

They tried the "brute force" method but also concluded it would be good to know why the CNN thinks it is a 1 GeV muon neutrino
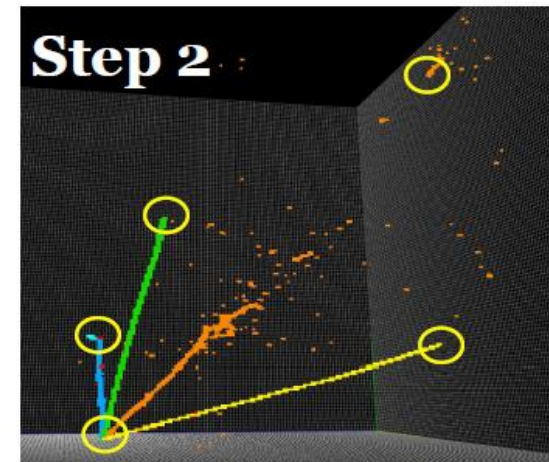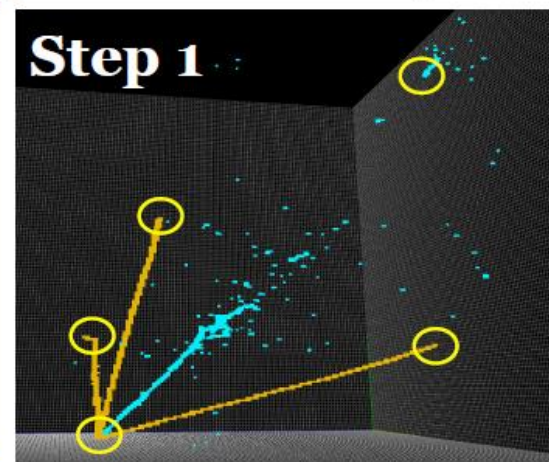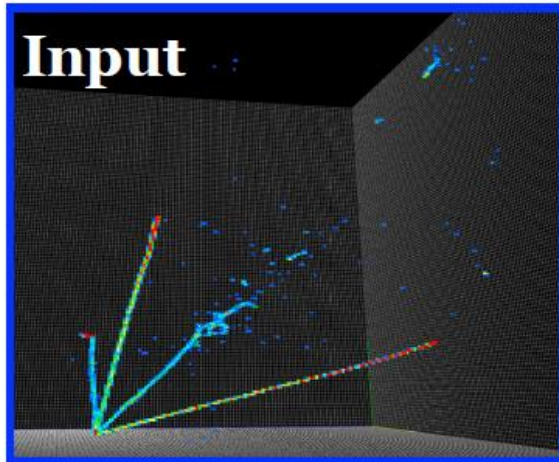
Data Reconstruction Chain

Extraction of hierarchical physics features...

1. Key points (particle start/end) + pixel feature extraction
2. Vertex finding + particle clustering
3. Particle type + energy/momentum
4. Interaction ("particle flow") reconstruction

Make it for 2D/3D data + the whole chain trainable

HK ML Discussions
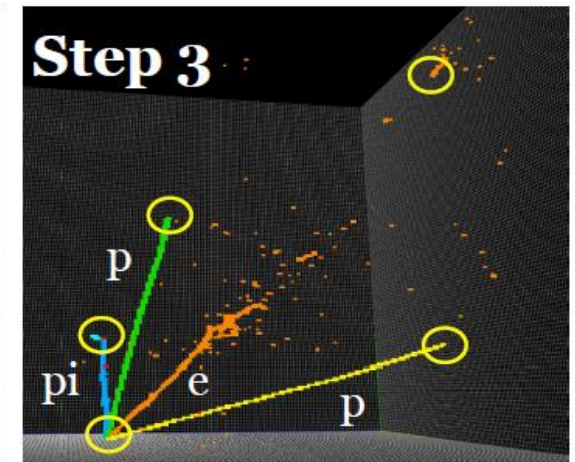
- Wrote last week to Patrick and Ishitsuka-san to mention the idea of the paraboloid reco but also the idea to have dedicated subtasks

- Patrick: *I agree it is good to divide into auxiliary subtasks (much like fiTQun already; or even [MicroBooNE/DUNE with full ML pipelines](#)); we have not yet had the people to progress on this path. It hasn't been tried yet and no arguments come to mind now, so would be good to explore.*

- Ishitsuka-san: *I have a similar understanding of ML for reconstruction. It is much easier and simpler to apply ML for a distinction of the cases, e.g. PID (e/mu/pi0/pi/gamma), number of rings and single/multi interactions. While reconstruction of vertex, direction, and momentum is more complicated with a large degree of freedom. We need some considerations to optimize the combination of ML and other methods.*

$\Rightarrow$ **We all seem to agree!**

$\Rightarrow$ **Possibly a mixture of classical algorithms and ML!**

$\Rightarrow$ **CNN are fine but depending on subtask other ML concepts might be used!**

Strategy

- Worth to try with brute force method: feed q,t to ResNet18

- But in parallel think about alternative approaches

- Important considerations:
  - What would be the required input information? E.g. vertex position …
  - Is it reasonable to assume that this information will be available?
  - What is the information we want to get out of the subtasks? A simple classification e.g. SV/MV? Number of vertices? Position of vertices?

- Then think what is the right concept: CNN? GNN? Classical concepts? Or???

- Makes the job more complicated but also (hopefully) more fun! More possibilities to come up with own ideas and explore them!

Some quick ideas

- Current 2D representation might not be ideal to find rings with CNN

- Ring might appear divided

- Assumption: known vertex

- Can we project PMT hits on a sphere around vertex?

- CNN to identify rings even when overlapping?

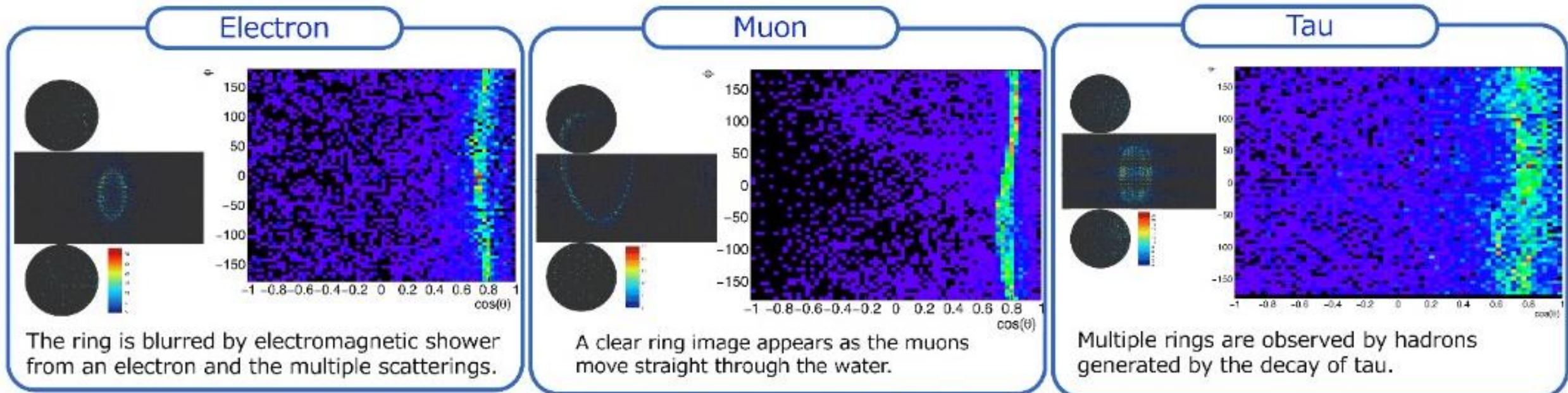- Might be even iterative process after finding ring hits, recalculate vertex position with better precision
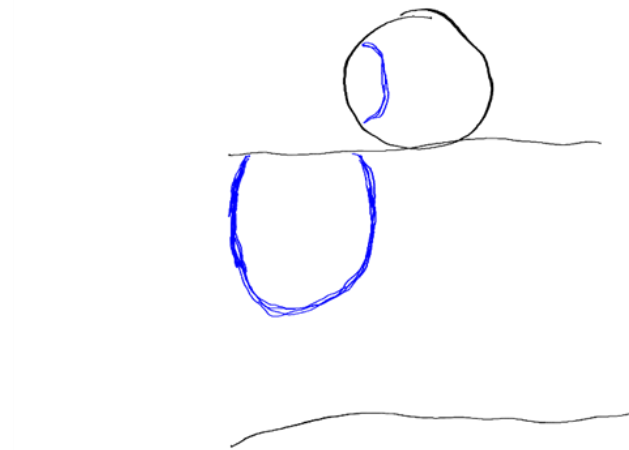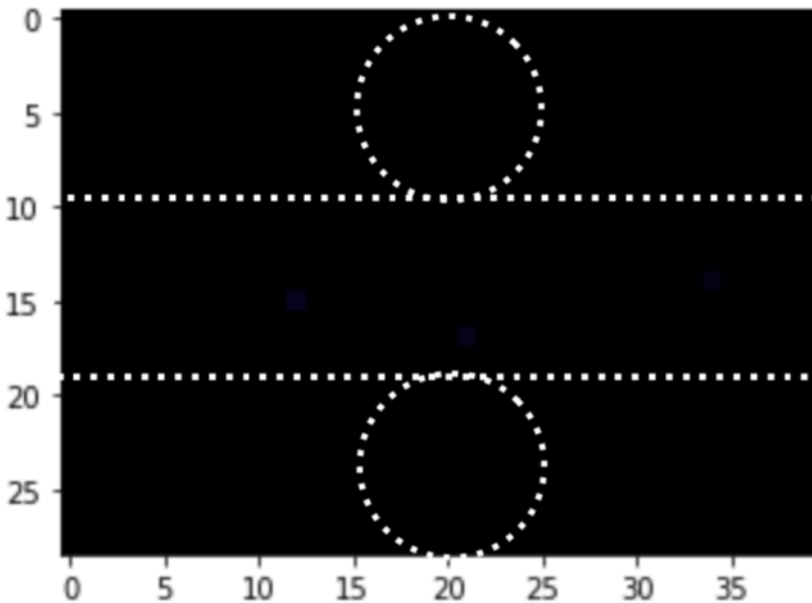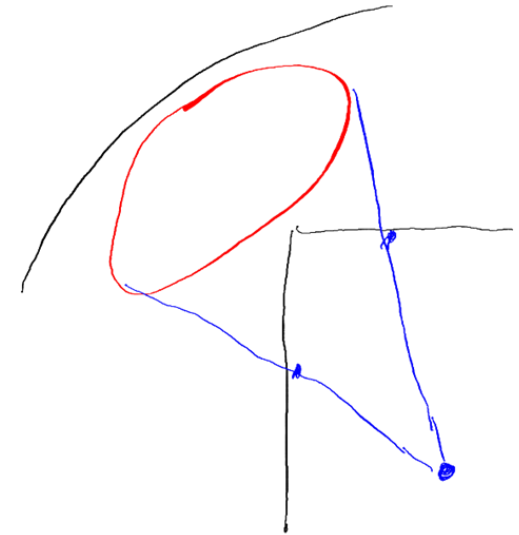
Full ring!
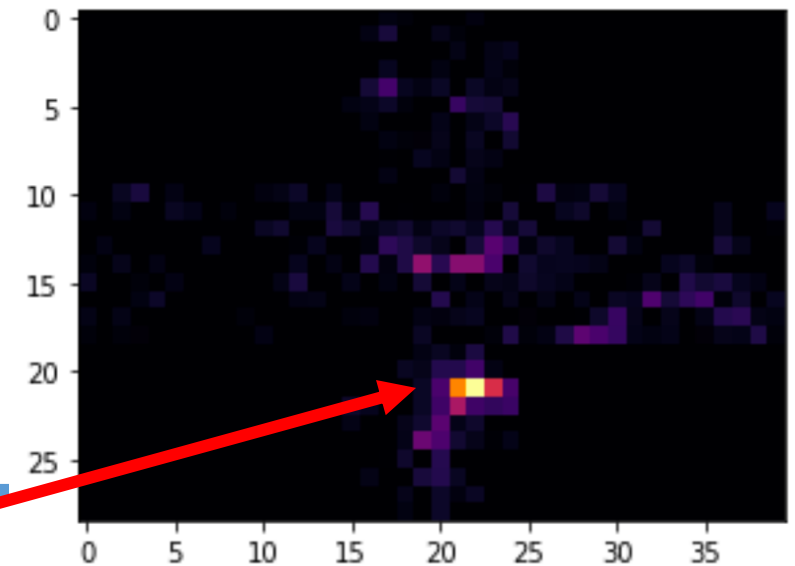
Some quick ideas

- Current 2D representation might work fine for some applications

- Particle Identification (PID) from "ring" shape should work

- Assumption: CNN does not really care about ring but local q distribution

- In this case, 2D projection should not be matter

- Good results achieved with HK ML

Electron — The ring is blurred by electromagnetic shower from an electron and the multiple scatterings.

Muon — A clear ring image appears as the muons move straight through the water.

Tau — Multiple rings are observed by hadrons generated by the decay of tau.

Some quick ideas

- Use GraphSage (GNN) to classify the hits with the information q,t,x,y,z for each PMT

- Could be done without correcting for vertex time or with correction

- Helpful to eliminate noise hits from images? Or ring hits?

- Assumption: ring found and particle identified

- Could we use CNN to determine precisely vertex position?

- Or use DBSCAN (cluster algorithm) to find segments and then connect segments to rings?

- Assumption: All clear ring hits identified and removed from hit list

- Could we use CNN to search for weak rings with not connected hits?
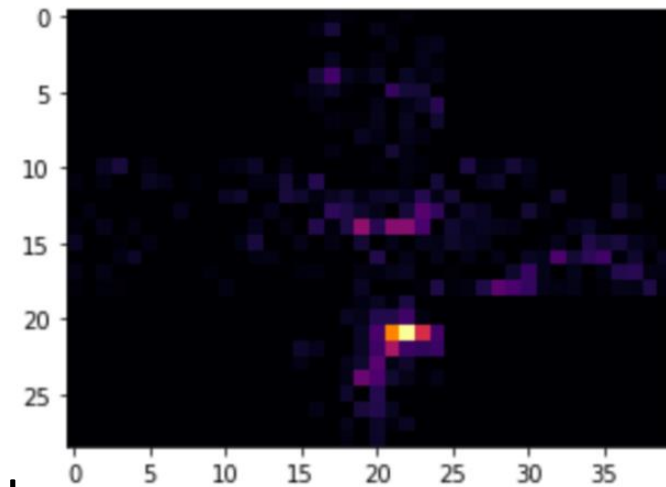
Conclusions

- Do not stop to run ResNet18 with your MC!
- Will serve as a good reference but idea is to move towards chain of reconstruction with several stages
- Will give better control on reconstruction and optimized architecture can be used for each step
- Good (or bad) news: Gives a lot of freedom to analyse and solve the problems (as long as it fits in the overall reco chain)
- I think it is more fun like this but yes, also more work!
- Some quick ideas are already there
- Important: Everyone can come up with own ideas!
- Can also explore ideas as Master thesis topics!

Pre-Vertex finder

Hit classifier

Ring finder

PID

Track vertex finder

Event vertex finder