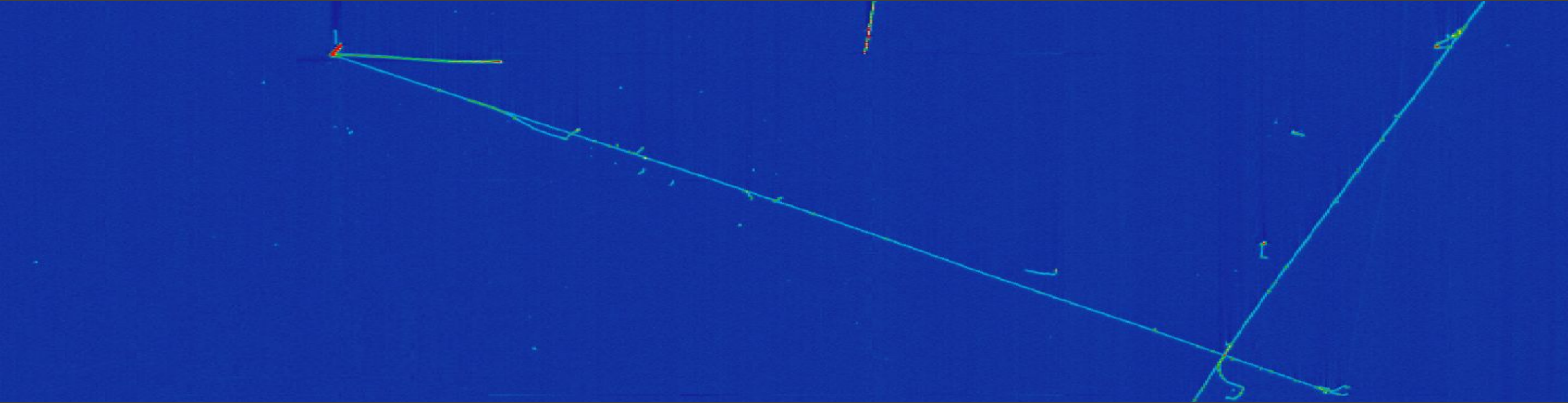


Machine Learning for Particle Imaging Detectors in Experimental Neutrino Physics

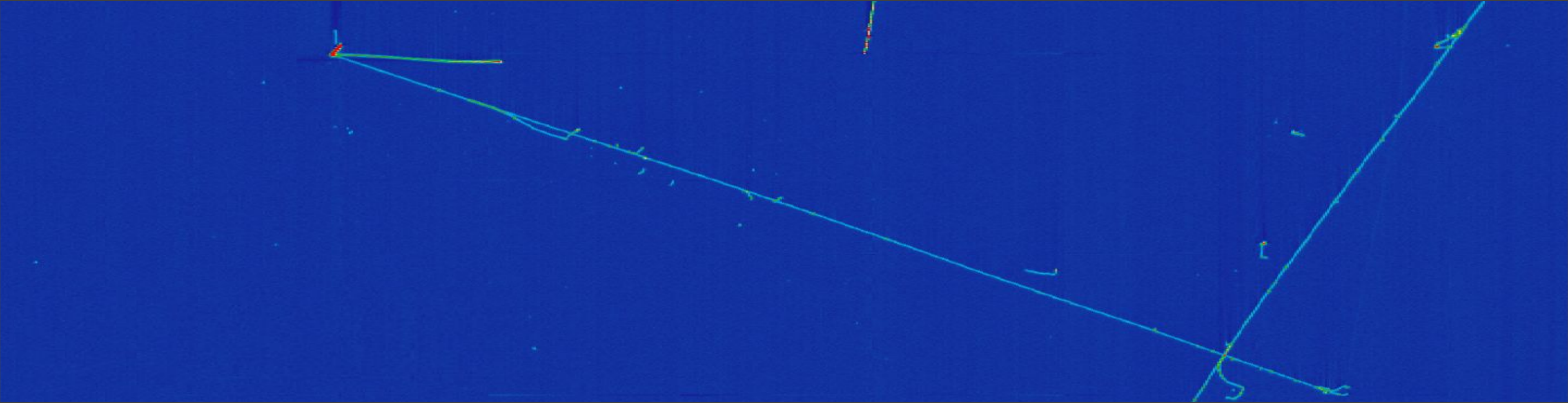
Kazuhiro Terao

SLAC National Accelerator Lab.

MLHEP @ DESY (July. 11th 2019)

U.S. DEPARTMENT OF ENERGY Office of Science

SLAC NATIONAL ACCELERATOR LABORATORY



Outline

1. Neutrino detectors
2. Machine Learning & Computer Vision Applications
3. ML-based Neutrino Data Reconstruction Chain
4. Summary

U.S. DEPARTMENT OF ENERGY
Office of Science

SLAC NATIONAL ACCELERATOR LABORATORY

Detectors for Neutrino Oscillation Experiments

Outline

1. **Neutrino detectors**
2. Machine Learning & Computer Vision Applications
3. ML-based LArTPC Data Reconstruction Ch
4. Summary

Neutrino Detectors: What's Important

Neutrino Oscillation Measurement

Use a neutrino source (flavour X), measure flavour Y at the detector

What's important?

Three important detector features for oscillation measurement

$$P(\nu_\mu \rightarrow \nu_e) = \sin^2 2\theta \sin^2\left(\frac{1.27\,\Delta m^2\, L}{E_\nu}\right)$$

Good Energy Resolution

Precise E_ν reduce oscillation uncertainty

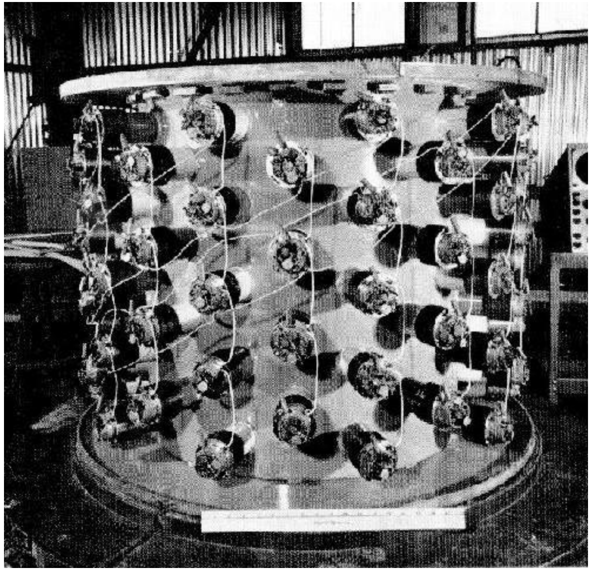
Large Mass (scalable)

"More" statistics to measure rare physics process

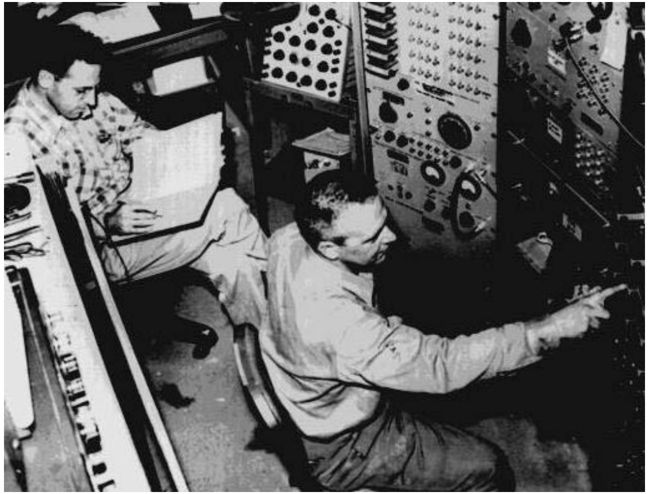
Particle ID Capability

Better ν identification background rejection

Neutrino Detectors: Early Days



Cd-doped water
0.4 ton, 100 PMTs
(1956)

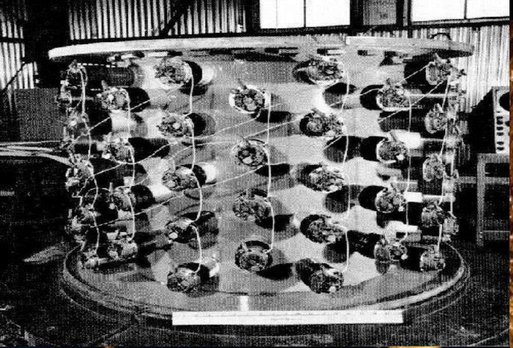


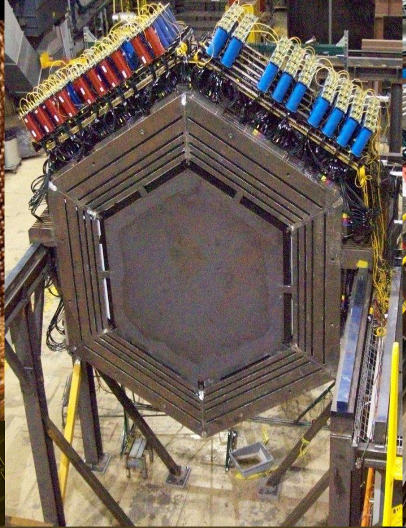
Inverse Beta Decay (IBD)

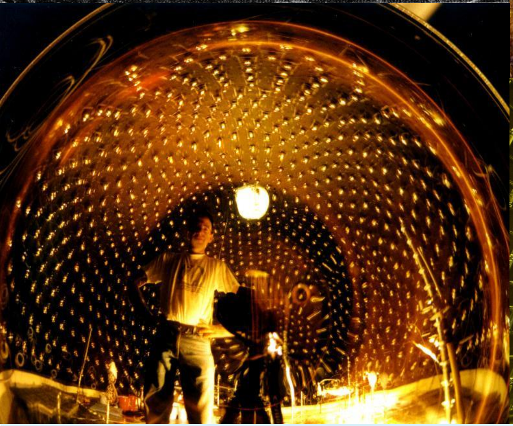
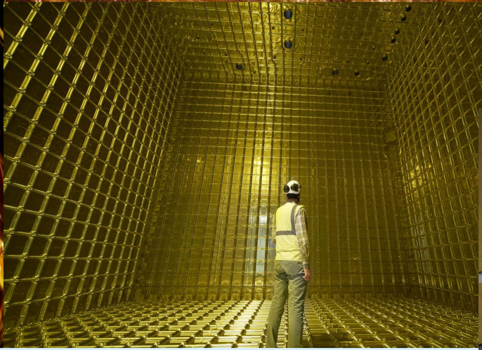
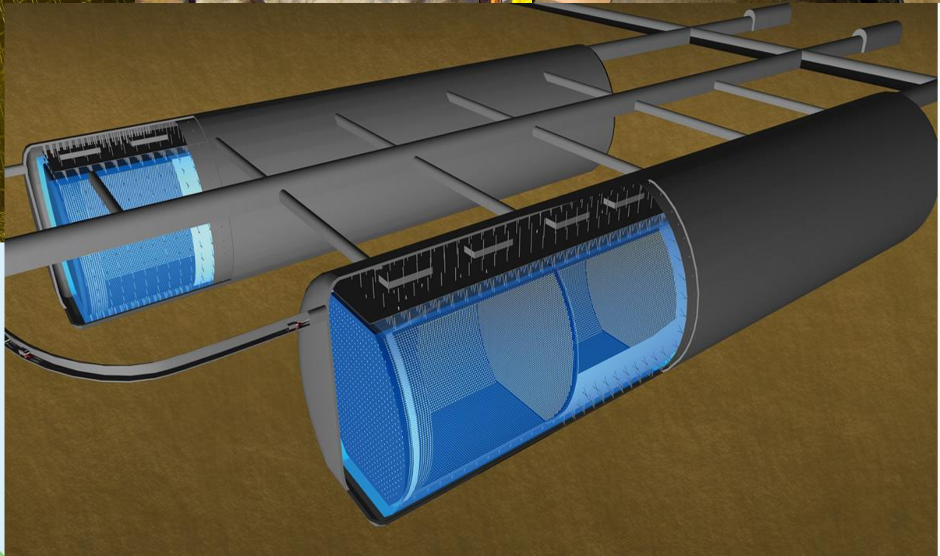
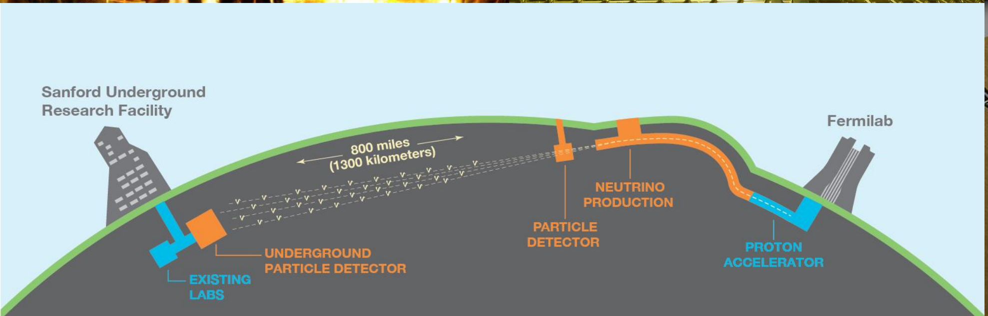
$$\bar{\nu}_e + p \rightarrow e^+ + n$$

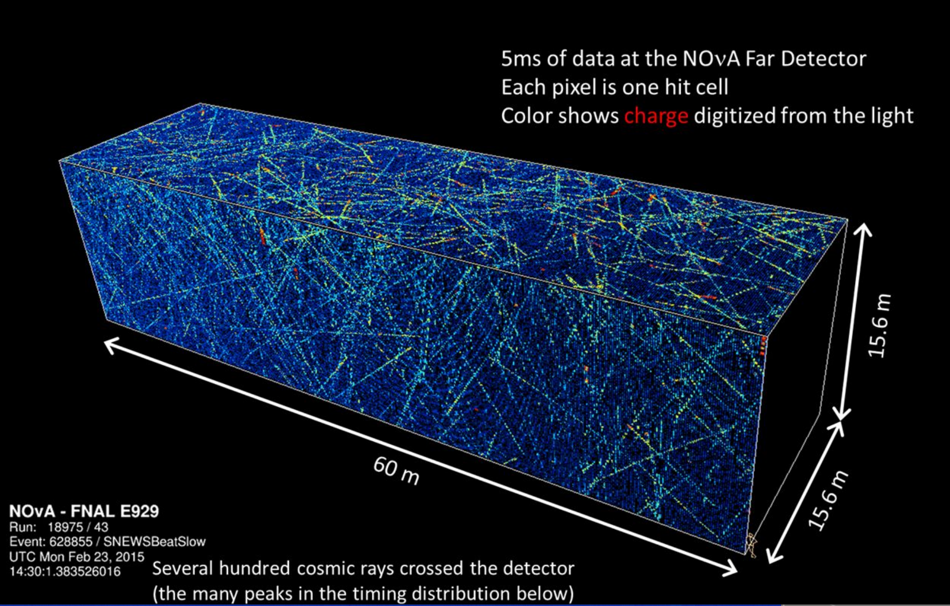
by Reines & Cowan (Nobel Prize 1995)

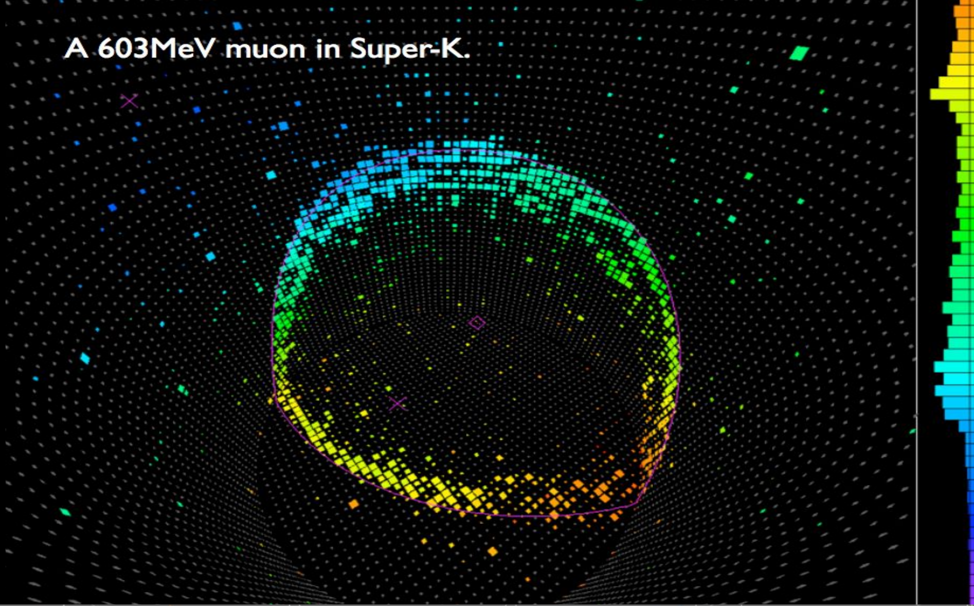
First neutrino detection

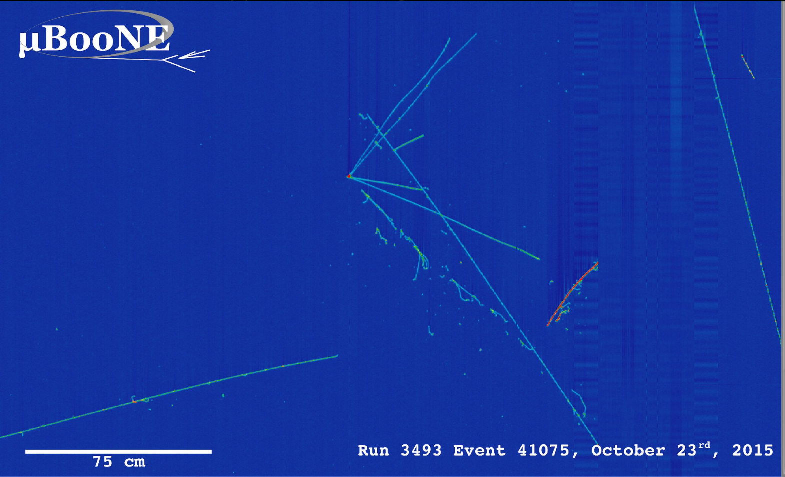


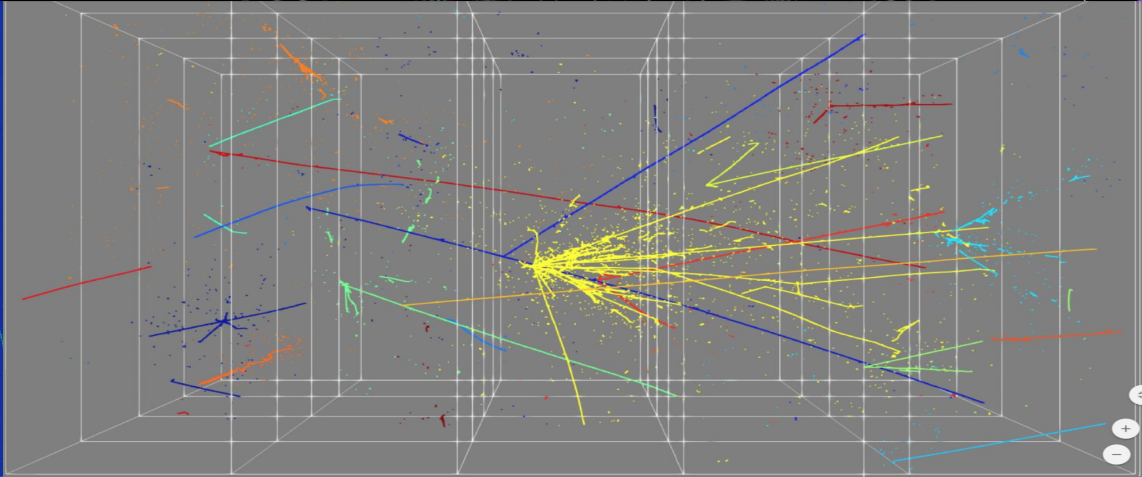


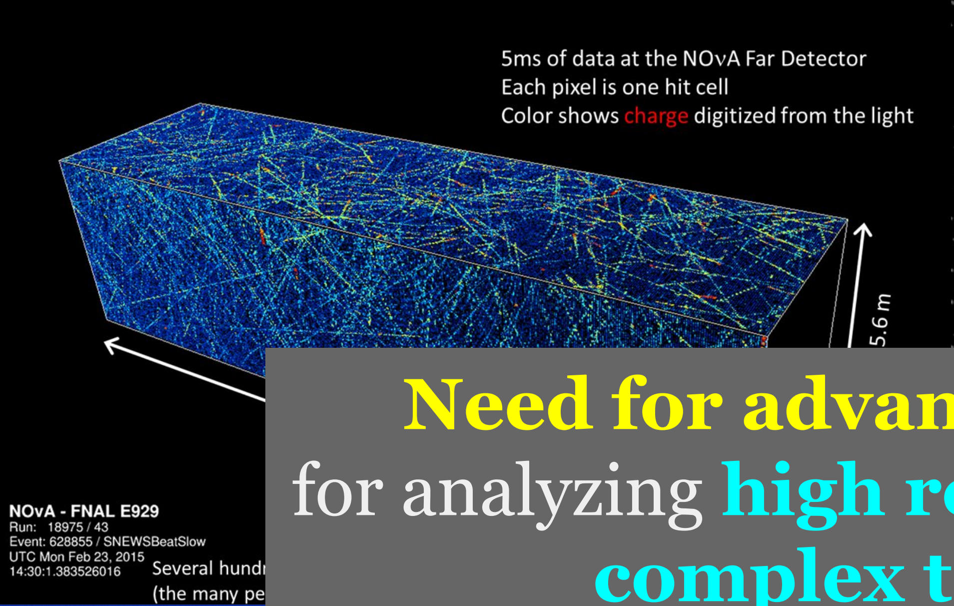




Sanford Underground
Research Facility
800 miles
(1300 kilometers)
Fermilab
NEUTRINO
PRODUCTION
PARTICLE
DETECTOR
UNDERGROUND
PARTICLE DETECTOR
PROTON
ACCELERATOR
EXISTING
LABS

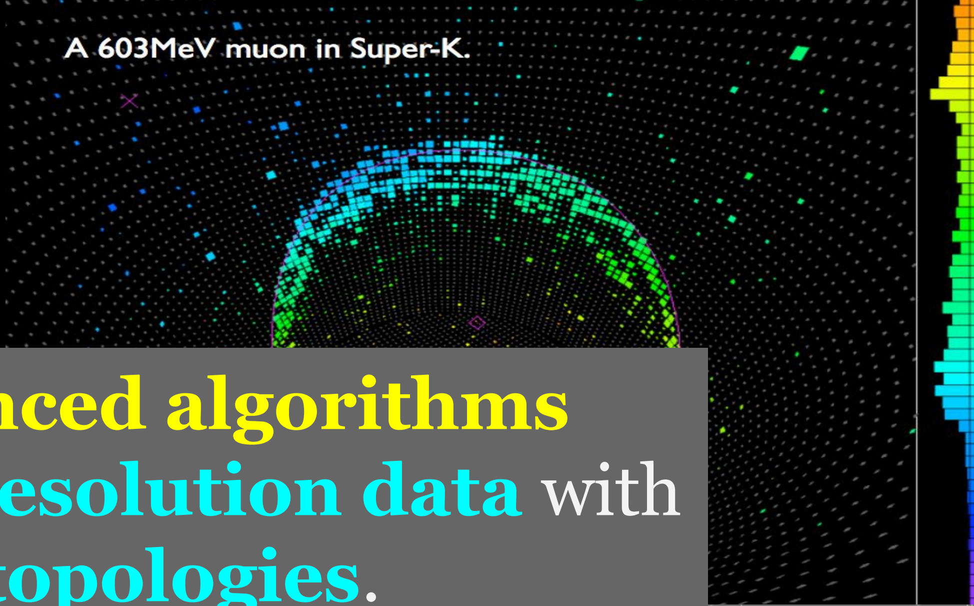

5ms of data at the NOνA Far Detector
Each pixel is one hit cell
Color shows charge digitized from the light
15.6 m
60 m
15.6 m
NOνA - FNAL E929
Run: 18975 / 43
Event: 628855 / SNEWSBeatSlow
UTC Mon Feb 23, 2015
14:30:1.383526016
Several hundred cosmic rays crossed the detector
(the many peaks in the timing distribution below)


A 603MeV muon in Super-K.

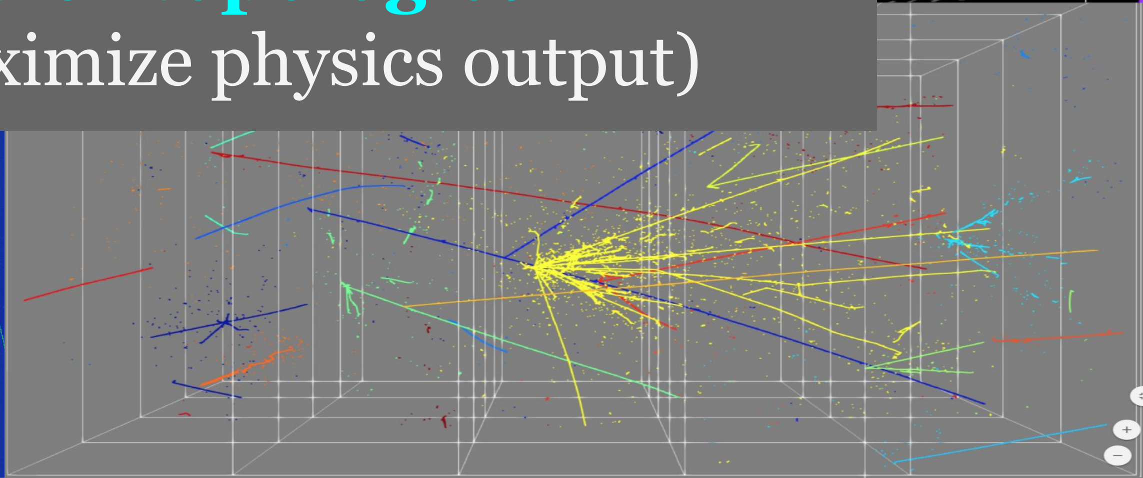

μBooNE
75 cm
Run 3493 Event 41075, October 23rd, 2015





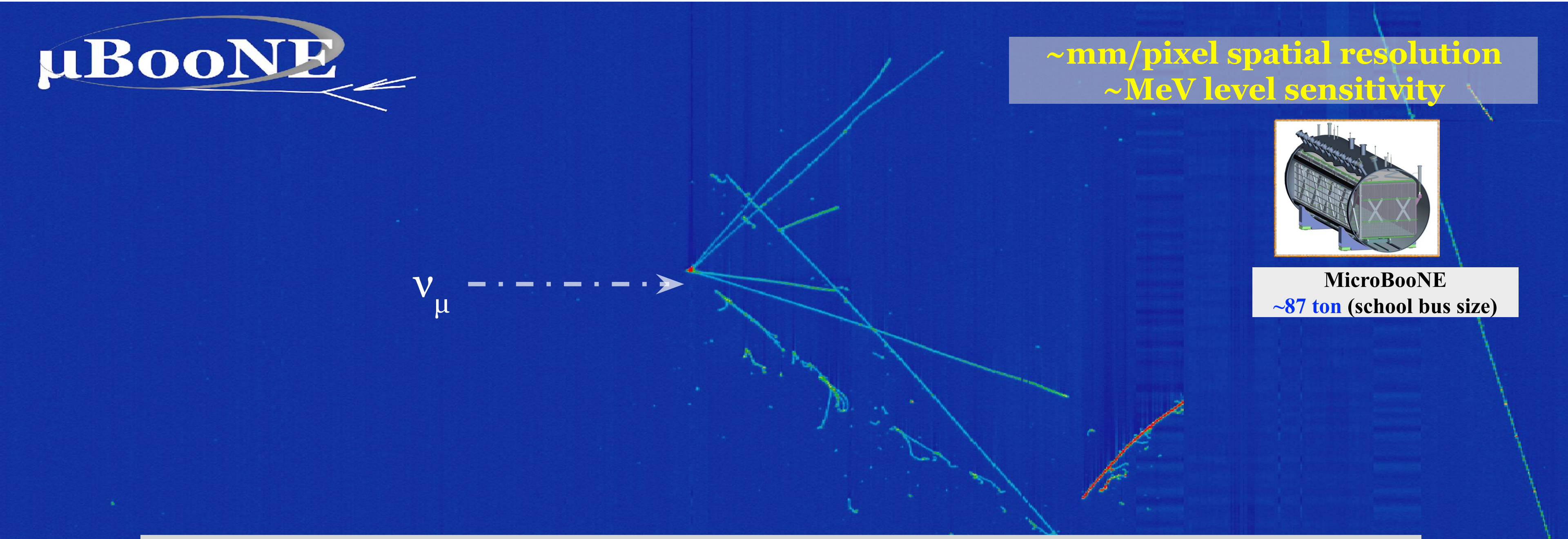


Need for advanced algorithms
for analyzing **high resolution data** with **complex topologies**.
(**goal**: maximize physics output)



Machine Learning & Computer Vision in Neutrino Physics

Time Projection Chambers



~mm/pixel spatial resolution
~MeV level sensitivity

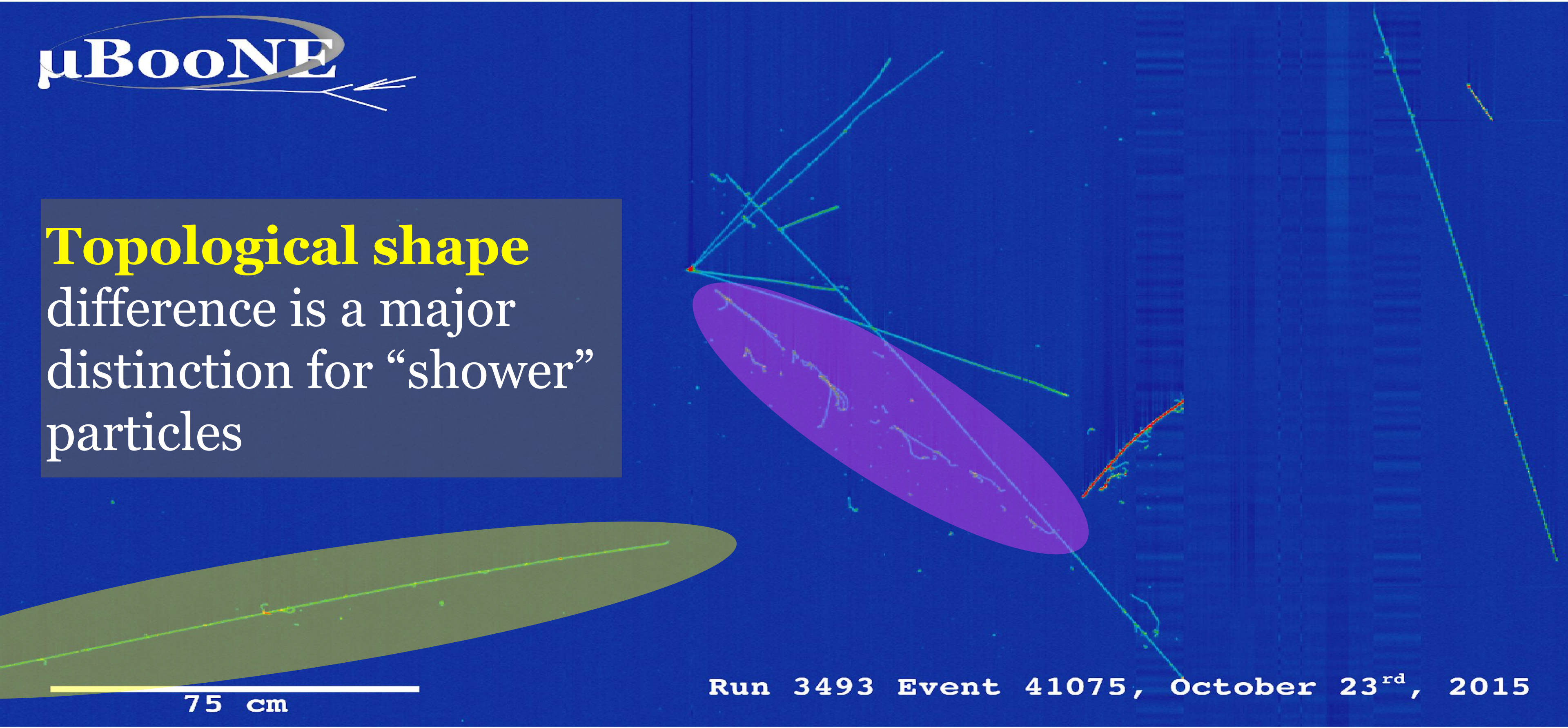
MicroBooNE
~87 ton (school bus size)

Liquid Argon Time Projection Chamber

- High resolution photograph of charged particle trajectories
- Calorimetric measurement + scalability to a large mass

Machine Learning & Computer Vision in Neutrino Physics
Time Projection Chambers

Topological shape difference is a major distinction for “shower” particles

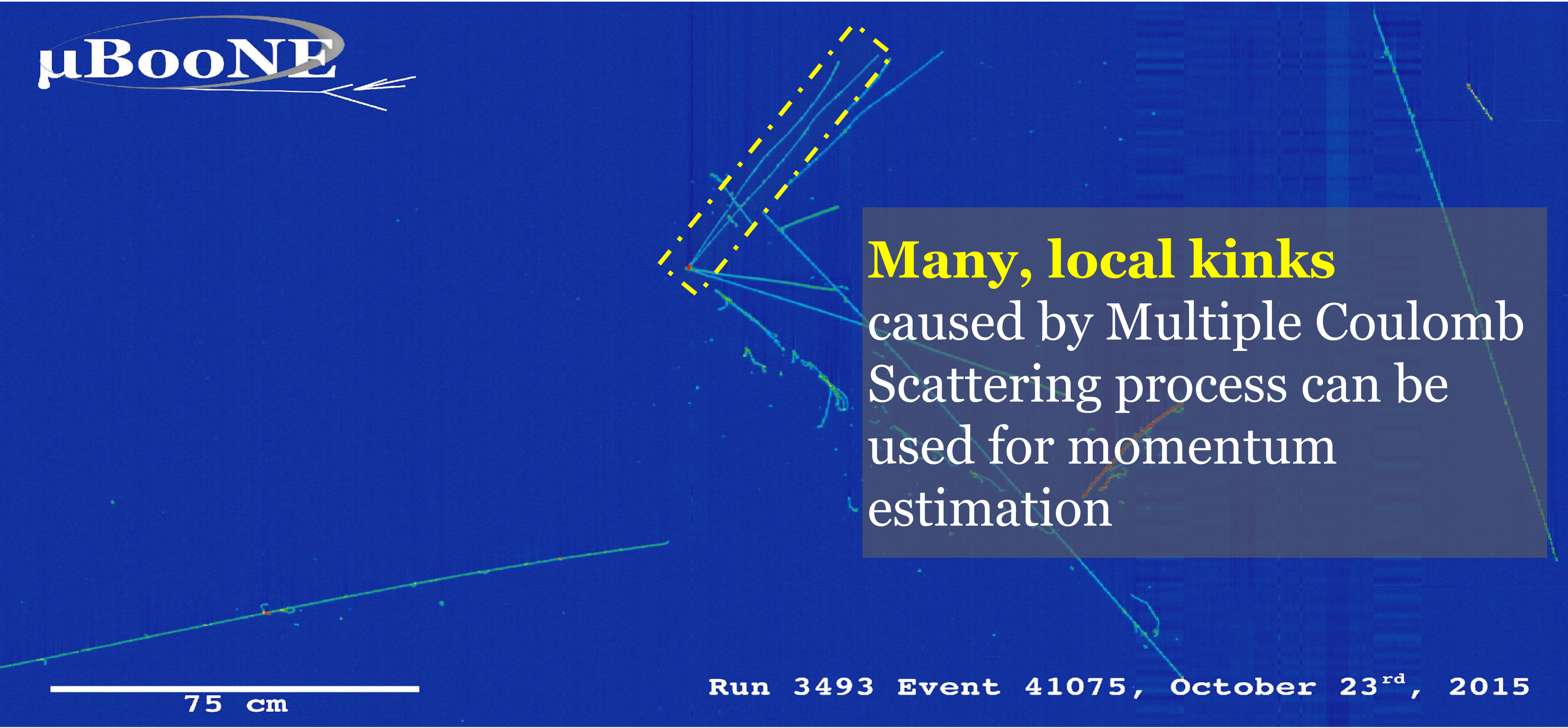


Time Projection Chambers

SLAC

Trajectory ends are distinct, and useful for seeding particle clustering and trajectory fitting

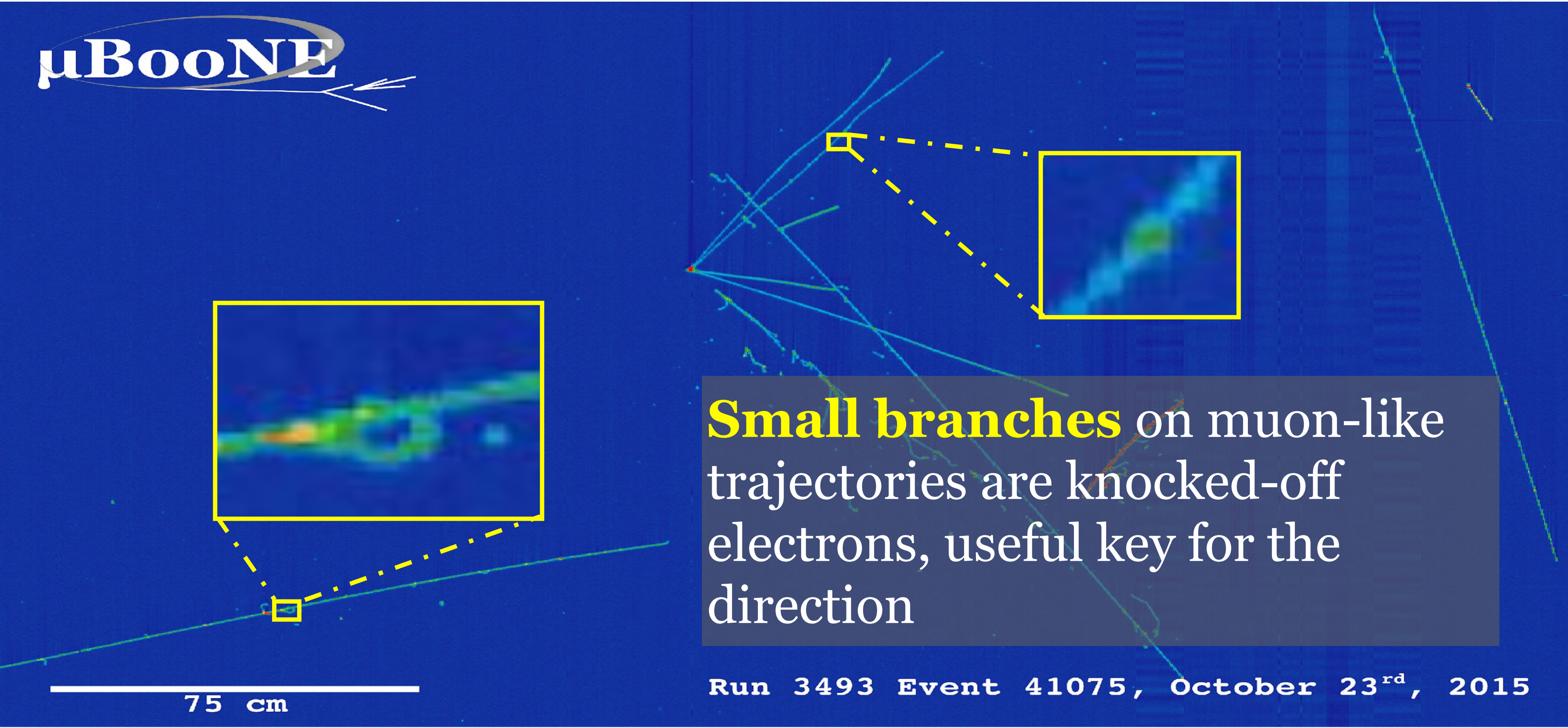
Time Projection Chambers



Many, local kinks
caused by Multiple Coulomb Scattering process can be used for momentum estimation

Machine Learning & Computer Vision in Neutrino Physics

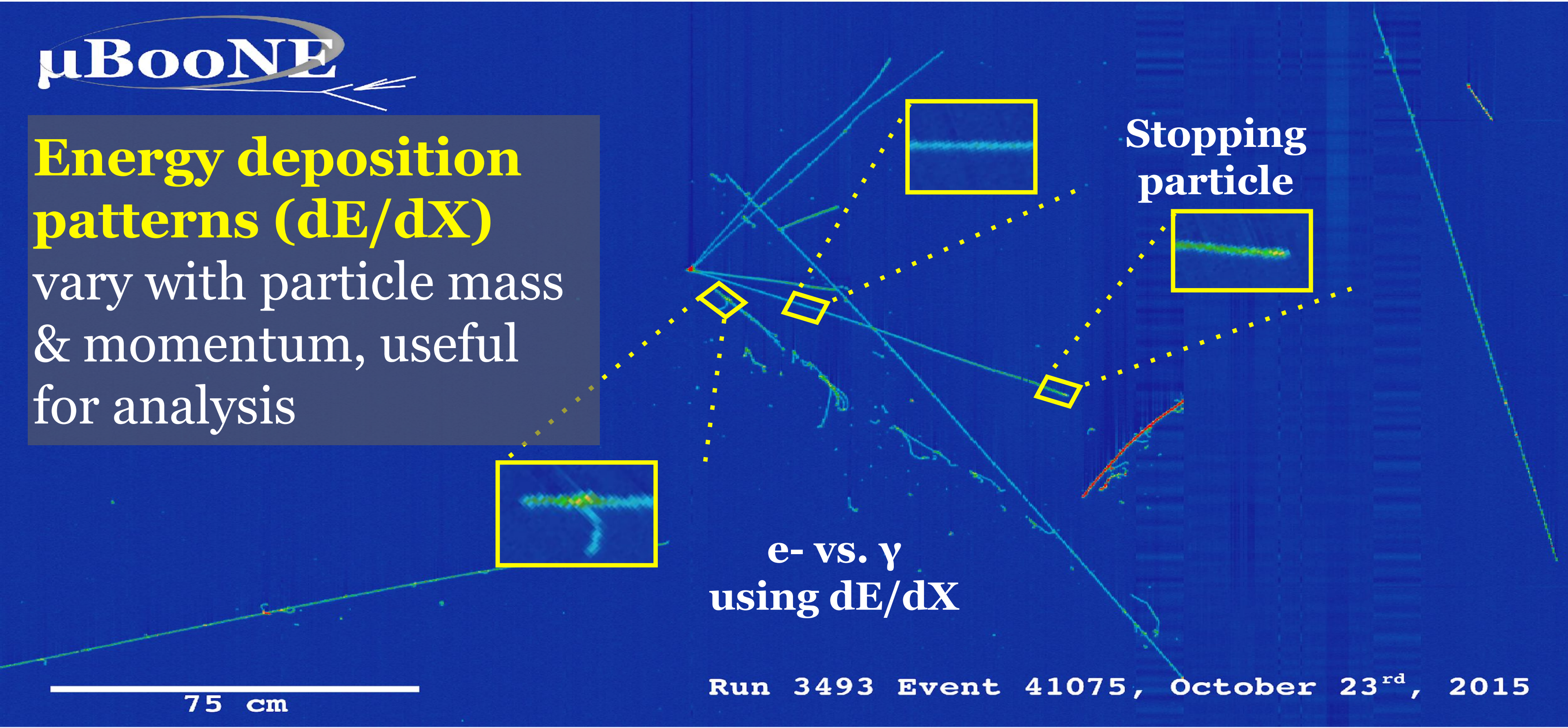
Time Projection Chambers



Machine Learning & Computer Vision in Neutrino Physics

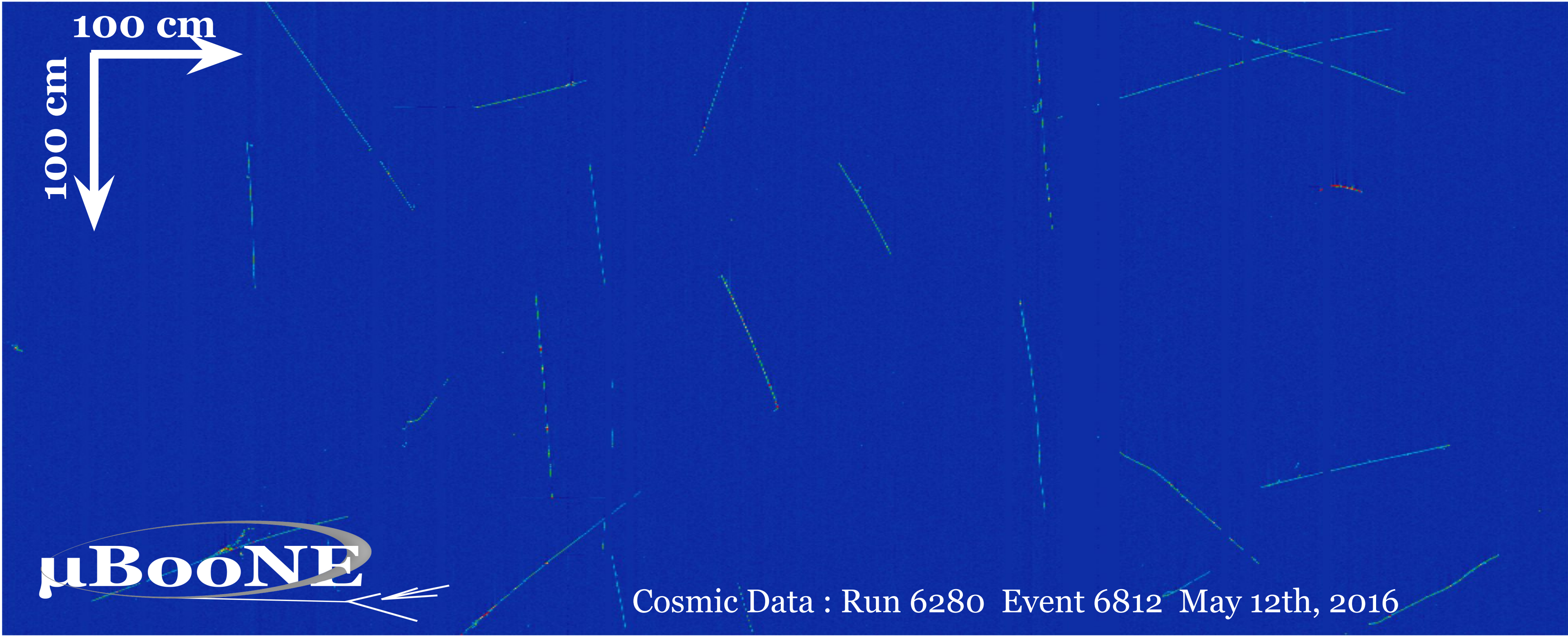
Time Projection Chambers

Energy deposition patterns (dE/dX) vary with particle mass & momentum, useful for analysis



Time Projection Chambers (slow ones)

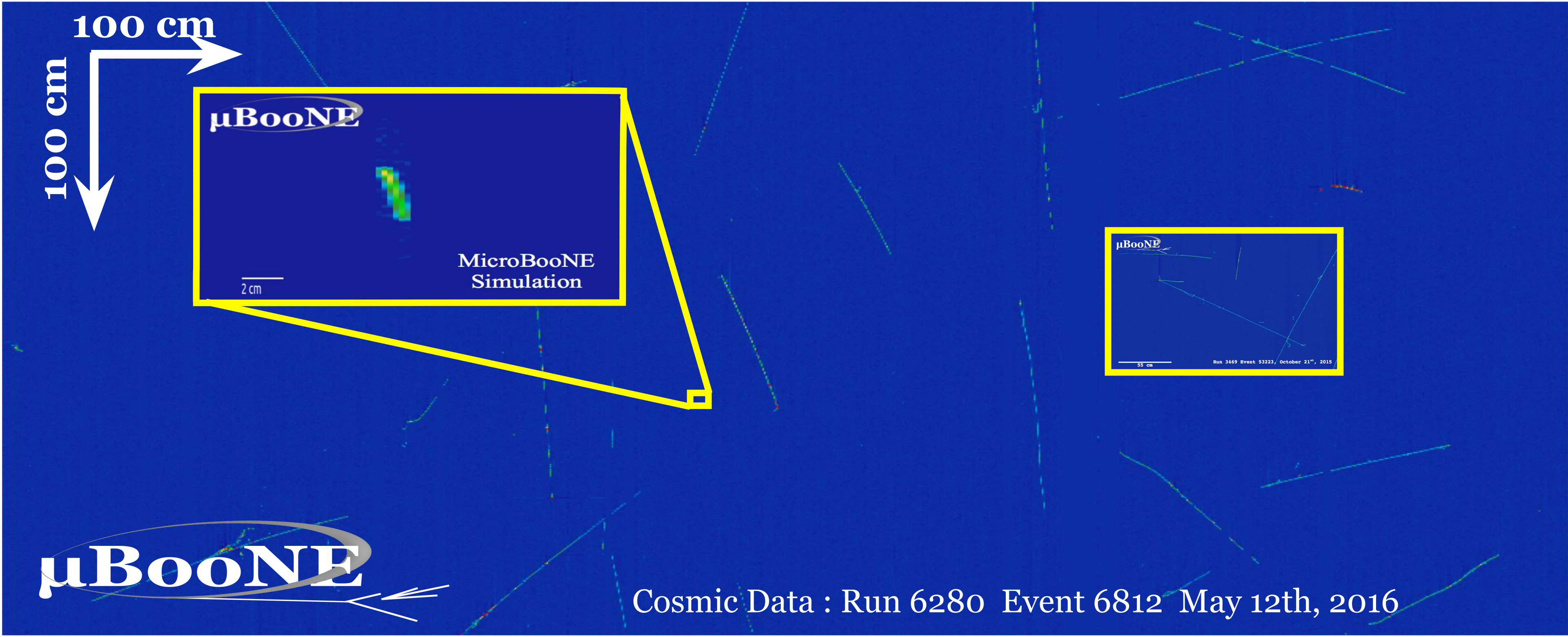
Do you see neutrino interaction here?



Machine Learning & Computer Vision in Neutrino Physics

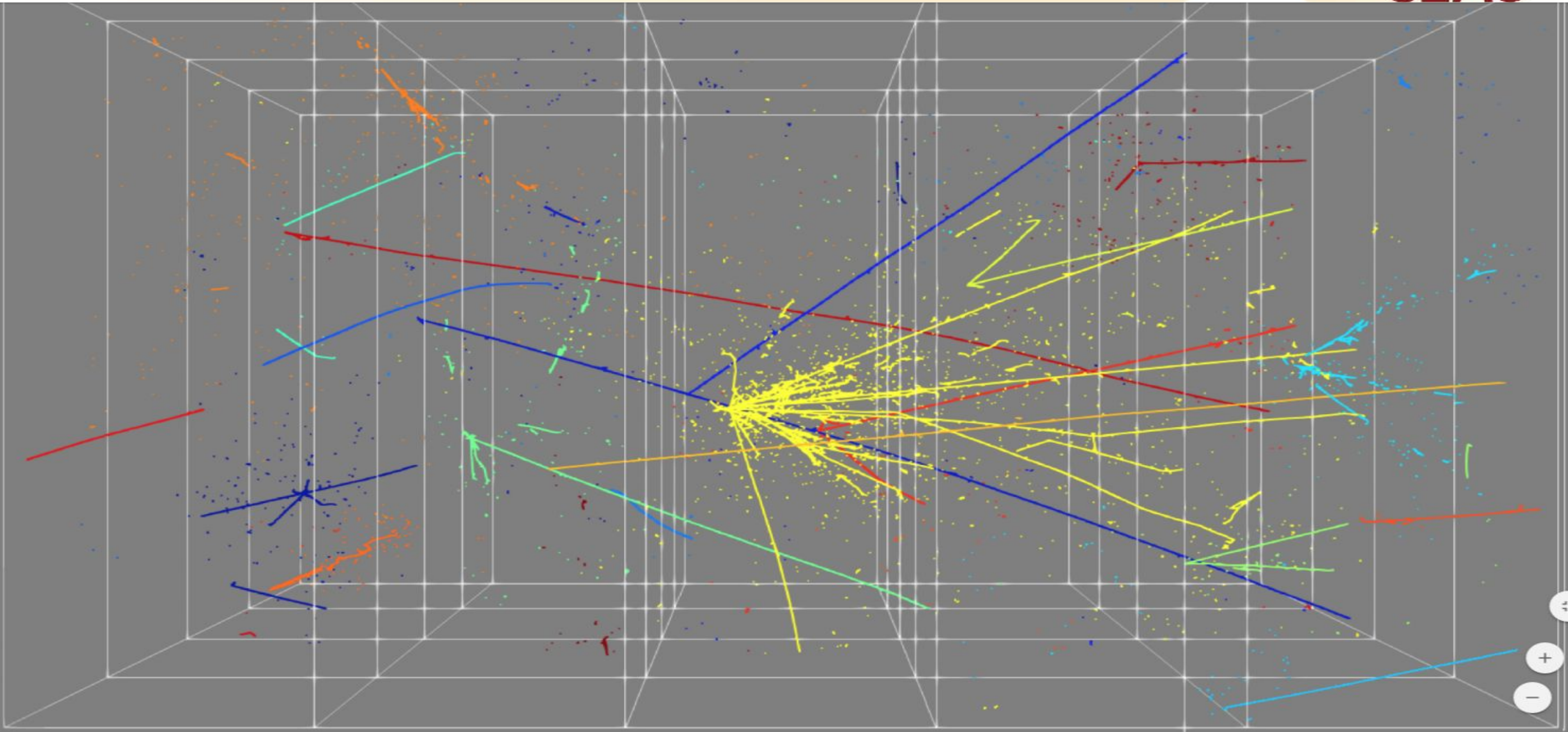
Time Projection Chambers (slow ones)

Now you do :)

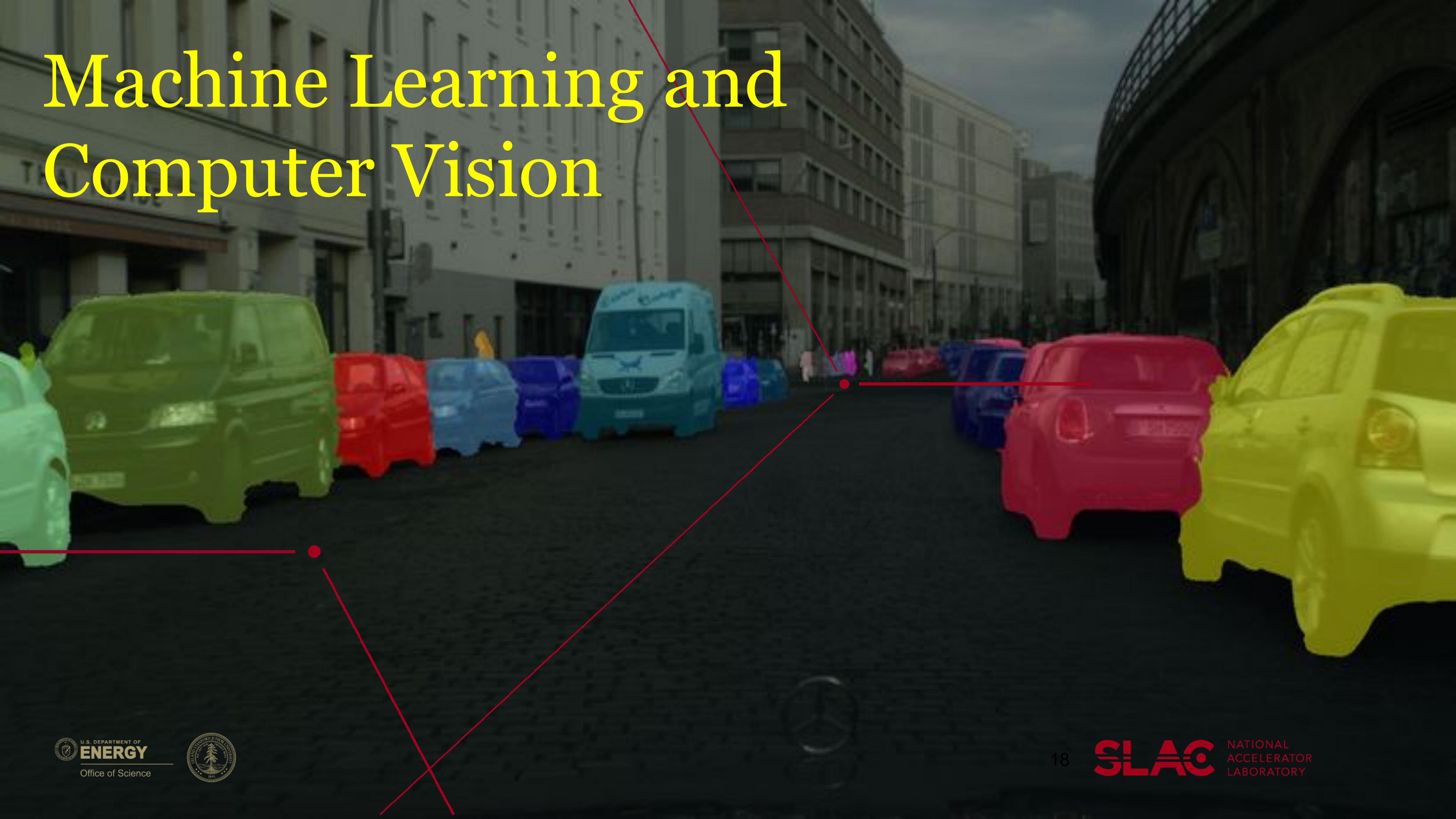


Machine Learning & Computer Vision in Neutrino Physics

Time Projection Chambers (3D ones)



Machine Learning and Computer Vision



You can find a cat? You can find a neutrino!

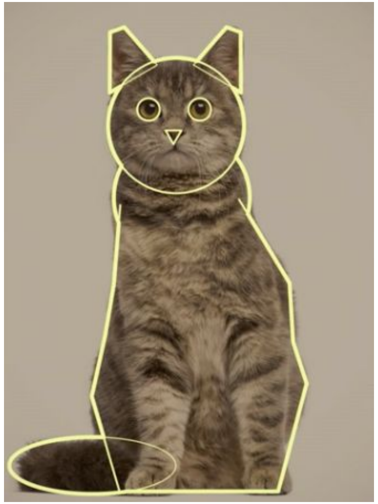
How to write an algorithm to identify a cat?

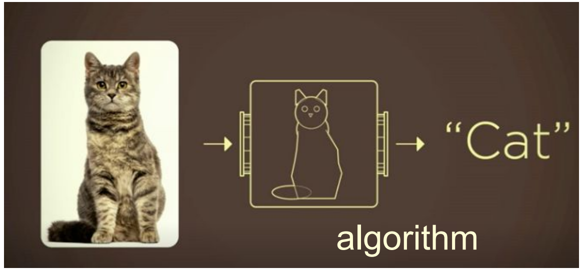
… very hard task …

16	08	67	15	83	09
37	52	77	23	22	74
35	42	48	72	85	27
68	36	43	54	21	33
79	60	10	25	54	71
18	55	38	73	50	47

Development Workflow for non-ML reconstruction

1. Write an algorithm based on physics principles





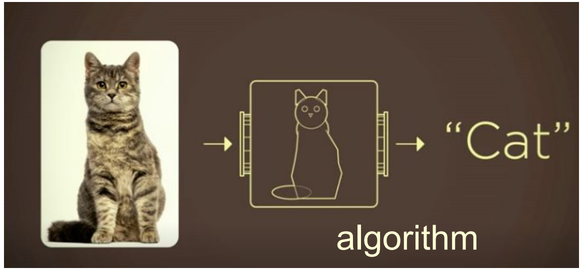
A cat = collection of certain shapes
(or, a neutrino)

Development Workflow for non-ML reconstruction

1. Write an algorithm based on physics principles
2. Run on simulation and data samples
3. Observe failure cases, implement fixes/heuristics
4. Iterate over 2 & 3 till a satisfactory level is achieved
5. Chain multiple algorithms as one algorithm, repeat 2, 3, and 4.

Partial cat
(escaping the detector)

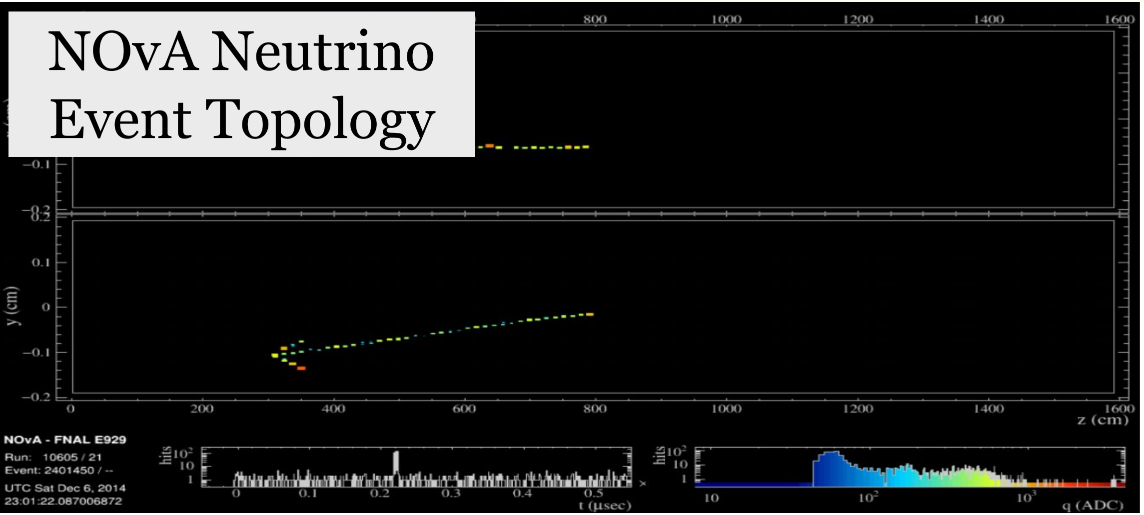
Stretching cat (Nuclear FSI)



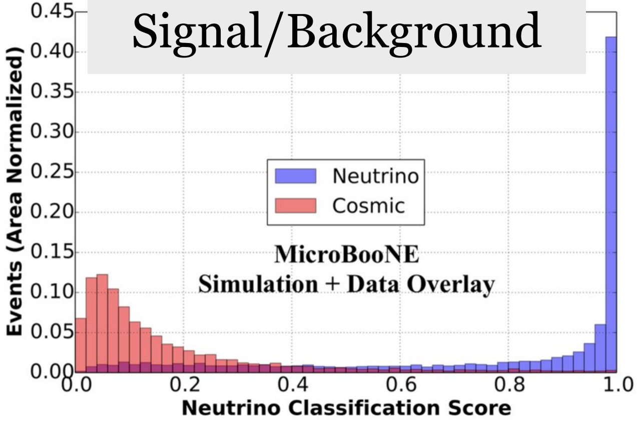
A cat = collection of certain shapes
(or, a neutrino)

Machine Learning & Computer Vision in Neutrino Physics
Image Classifications: a lot of applications

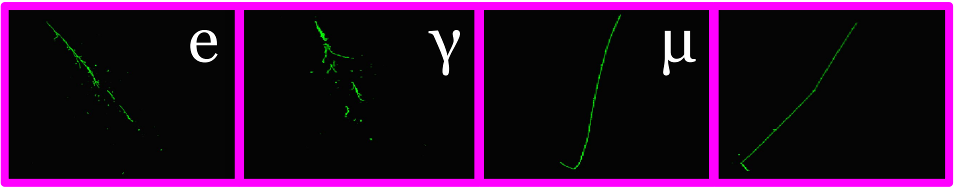
NOvA Neutrino Event Topology



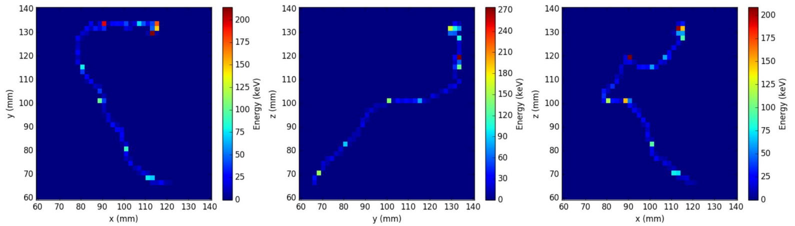
MicroBooNE Signal/Background



LArLIAT Particle Type Identification



NEXT Signal vs. Background



Machine Learning & Computer Vision in Neutrino Physics
Object Detection & Semantic Segmentation

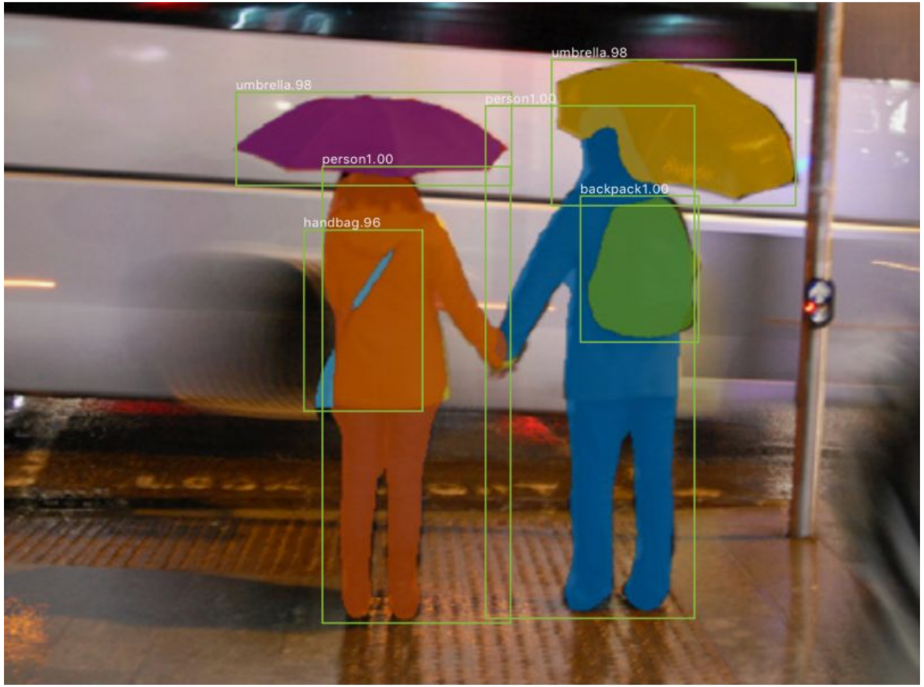


Image Context Identification

Hierarchy and Correlation of Context

"girl in pink dress is jumping in air."

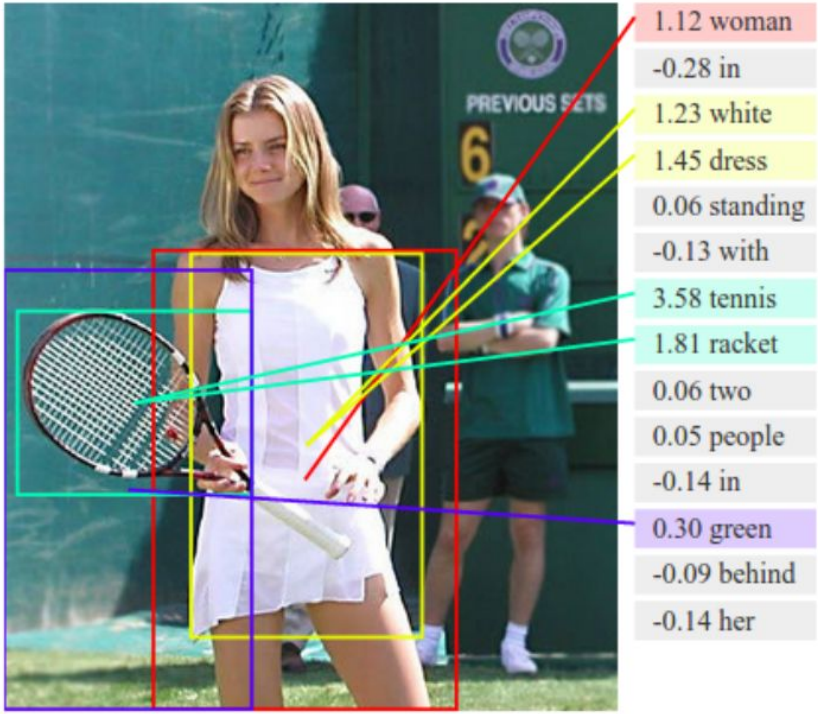
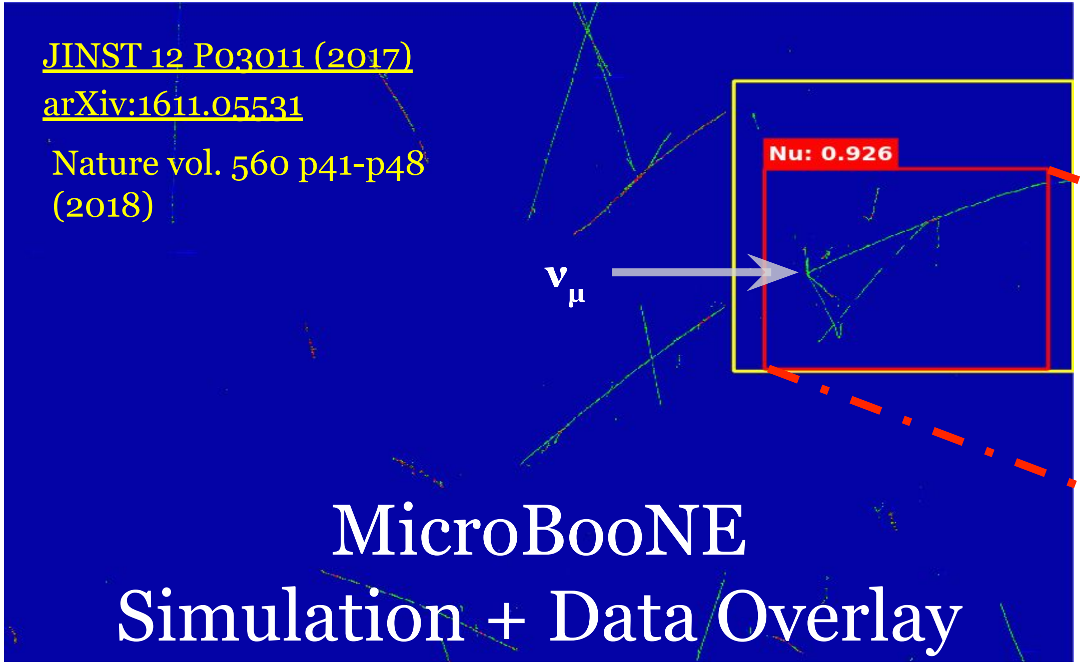
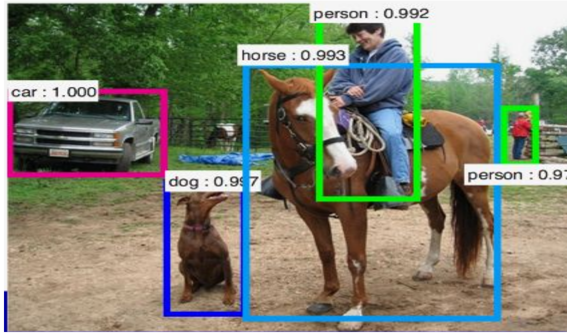
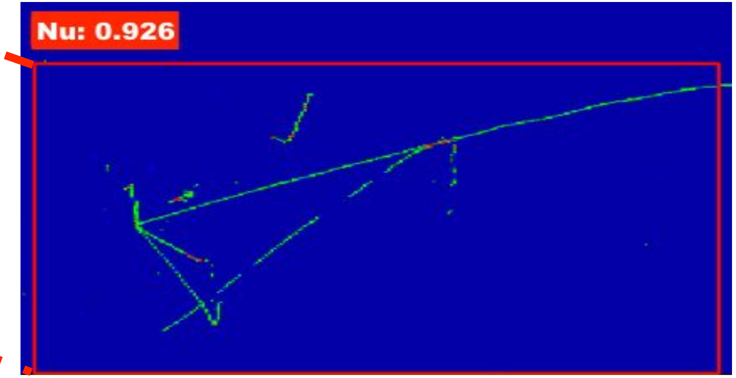


Image Context Correlation/Hierarchy Analysis

Neutrino Detection w/ R-CNN
(MicroBooNE LArTPC)





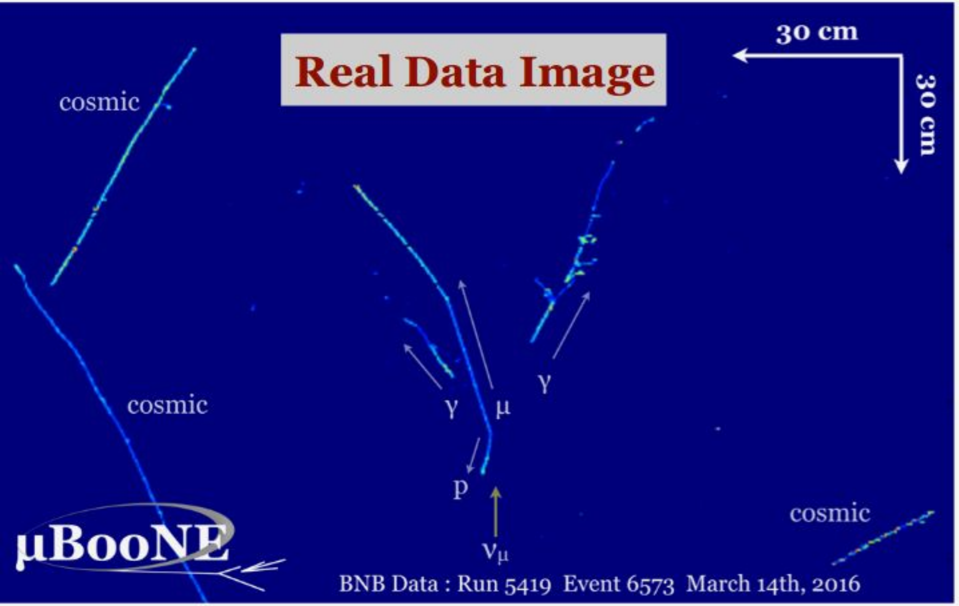


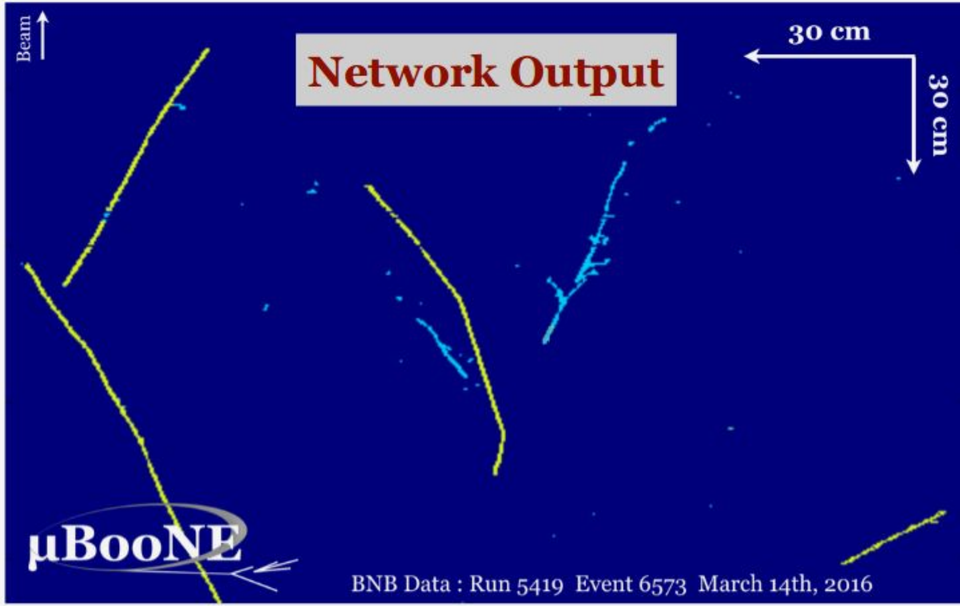
Task: propose a rectangular box that contains neutrino interaction
(location & size)

Machine Learning & Computer Vision in Neutrino Physics

Semantic Segmentation for Pixel-level Particle ID

Separate electron/positron energy depositions from other types at raw waveform level.
Helps the downstream clustering algorithms (data/sim comparison @ arxiv:1808.07269)



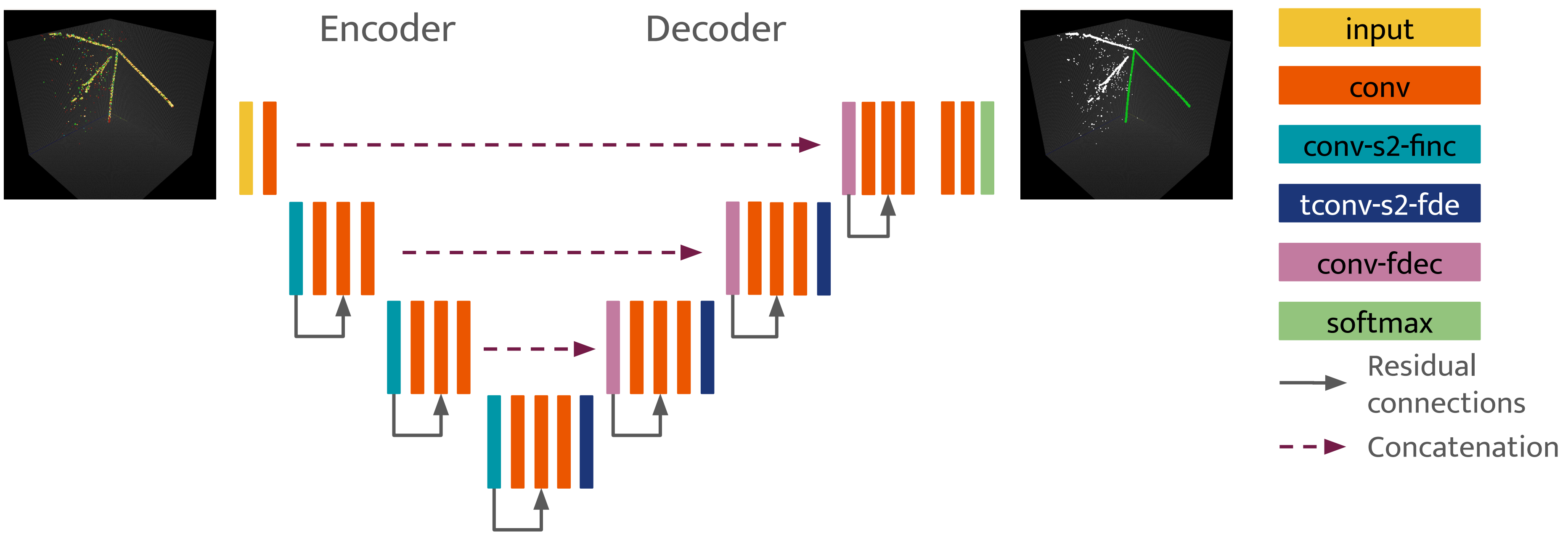


Network Input

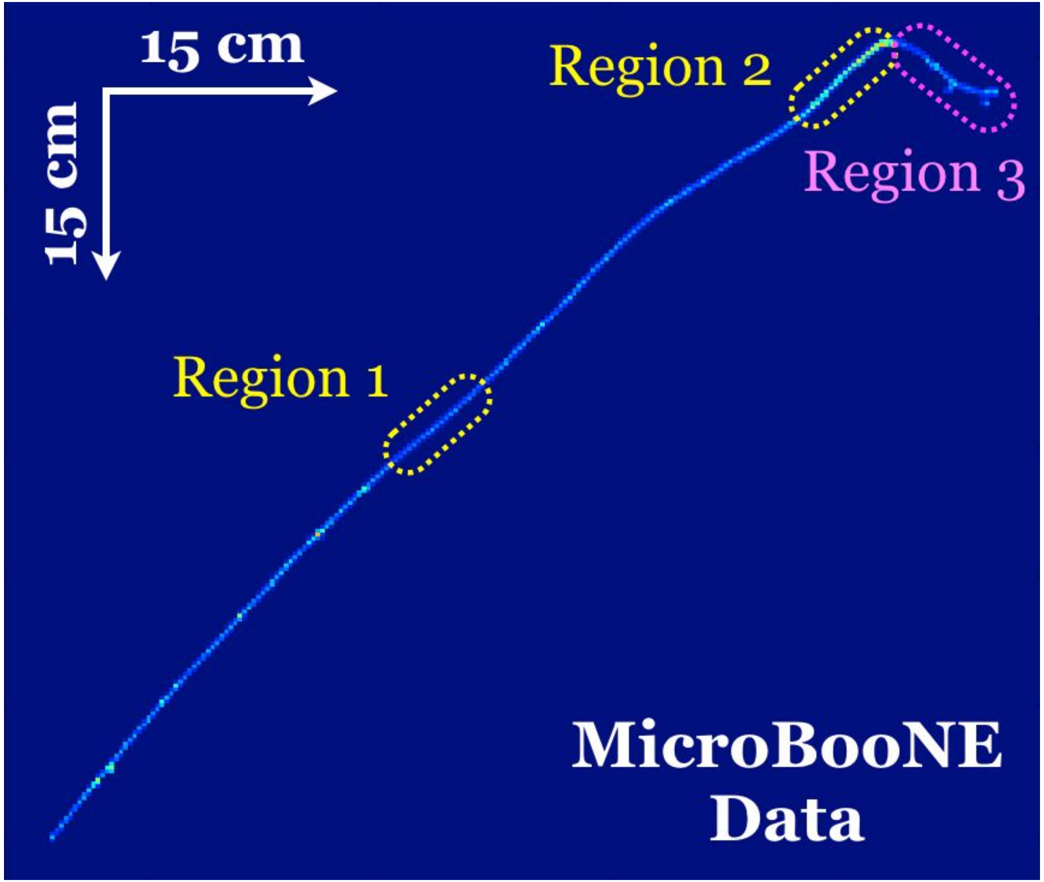
Network Output

Semantic Segmentation for Pixel-level Particle ID

Architecture: **U-Net + Residual Connections**

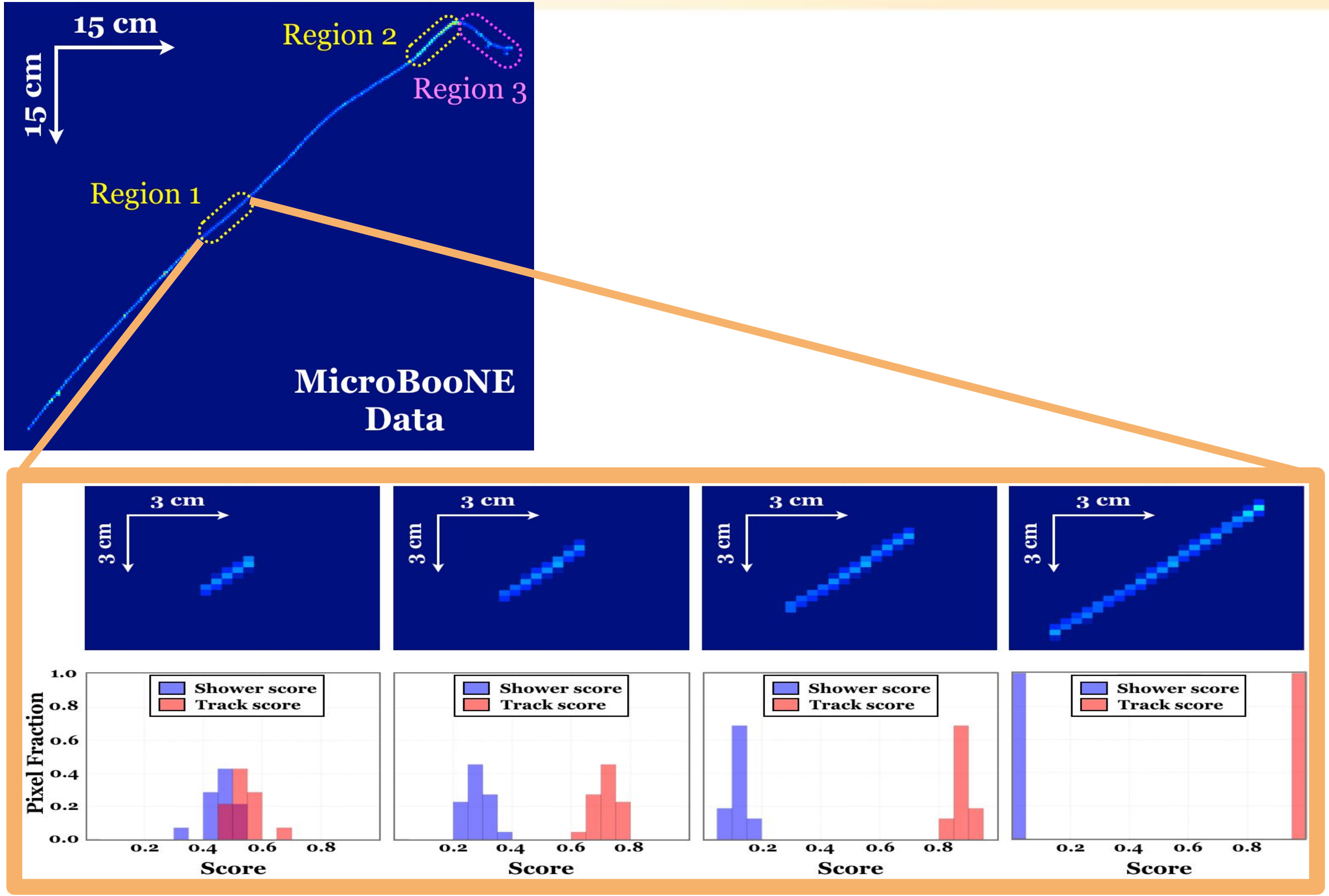


Fun Playing with Semantic Segmentation



Machine Learning & Computer Vision in Neutrino Physics

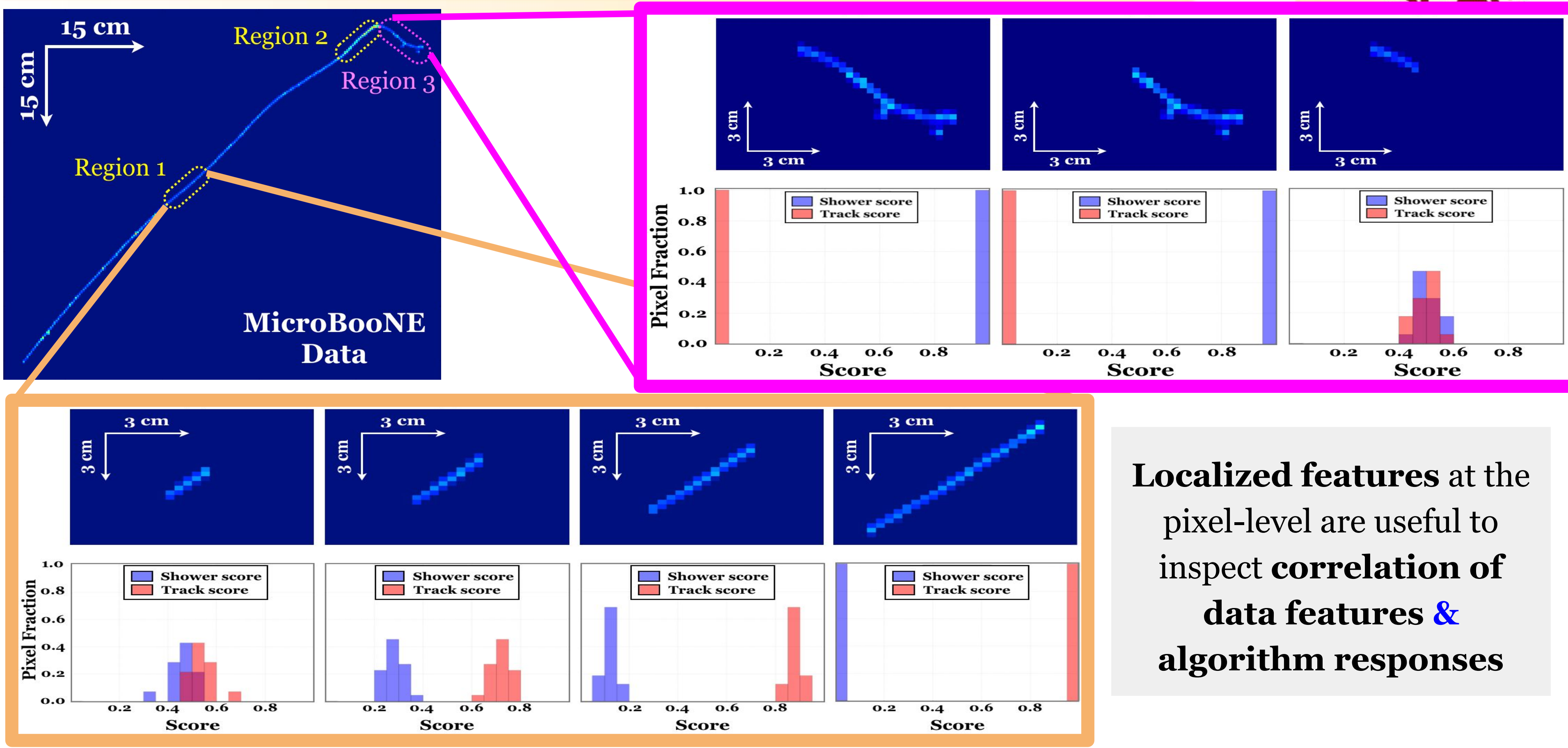
Fun Playing with Semantic Segmentation



Localized features at the pixel-level are useful to inspect **correlation of data features & algorithm responses**

Machine Learning & Computer Vision in Neutrino Physics

Fun Playing with Semantic Segmentation



Localized features at the pixel-level are useful to inspect **correlation of data features & algorithm responses**

Scalable CNN for Sparse Particle Imaging Data

"Applying CNN" is simple, but **is it scalable?**

LArTPC data is generally sparse, but locally dense

CNN applies **dense matrix operations**

In photographs, **all pixels are meaningful**

Figures/Texts: courtesy of Laura Domine @ Stanford

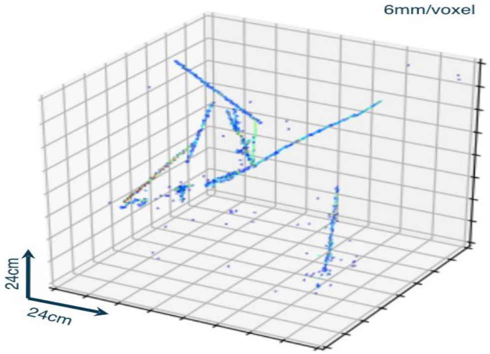
Scalable CNN for Sparse Particle Imaging Data

"Applying CNN" is simple, but **is it scalable?**

LArTPC data is generally sparse, but locally dense

CNN applies **dense matrix operations**

In photographs, **all pixels are meaningful**



<1% of pixels are non-zero in LArTPC data

Zero pixels are meaningless!

Figures/Texts: courtesy of Laura Domine @ Stanford

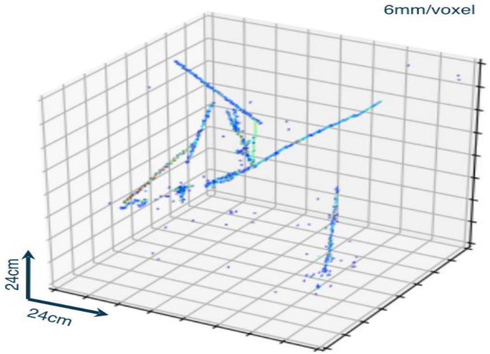
Scalable CNN for Sparse Particle Imaging Data

“Applying CNN” is simple, but **is it scalable?**

LArTPC data is generally sparse, but locally dense

CNN applies **dense matrix operations**

In photographs, **all pixels are meaningful**



<1% of pixels are non-zero in LArTPC data

Zero pixels are meaningless!

Figures/Texts: courtesy of Laura Domine @ Stanford

- **Scalability for larger detectors**
  - Computation cost increases linearly with the volume
  - But the number of non-zero pixles does not

Submanifold Sparse Convolutions

Many possible definitions and implementations of '*sparse convolutions*'...

Submanifold Sparse Convolutions (arxiv:1711.10275, CVPR2018):
https://github.com/facebookresearch/SparseConvNet

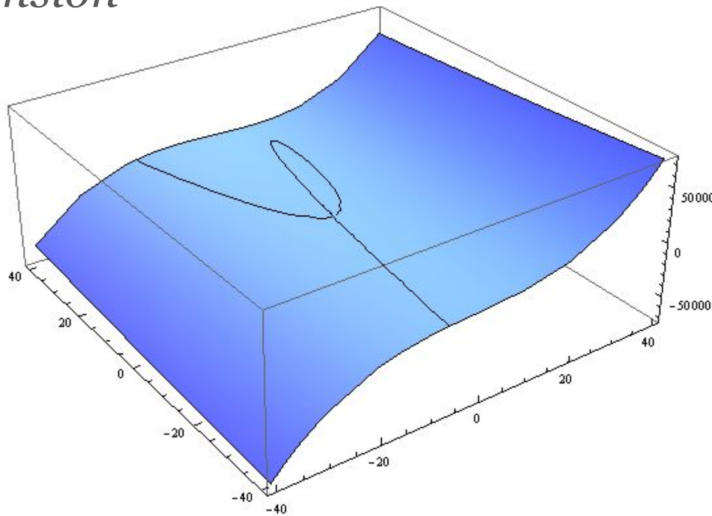
State-of-the-art on ShapeNet challenge (3D part segmentation)

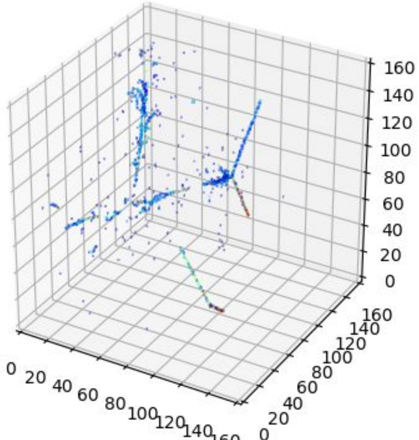
Submanifold Sparse Convolutions

Submanifold = *"input data with lower effective dimension than the space in which it lives"*

Ex: 1D curve in 2+D space, 2D surface in 3+D space

Our case: the worst! **1D curve in 3D space**...





Submanifold Sparse Convolutions

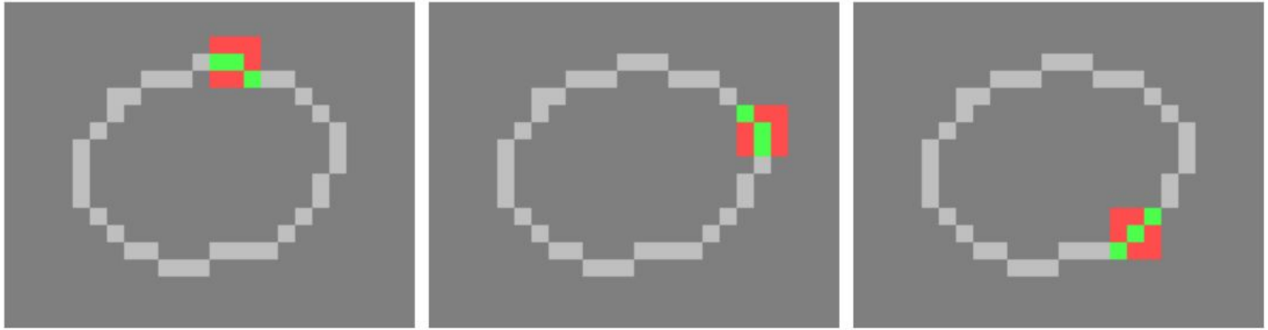
1. **Resources waste** of dense convolutions on sparse data
2. **Dilation problem**

- 1 nonzero site leads to 3^d nonzero sites after 1 convolution
- How to keep the same level of sparsity throughout the network?

[3D Semantic Segmentation with Submanifold Sparse Convolutional Networks](#) (arxiv: 1711.10275)

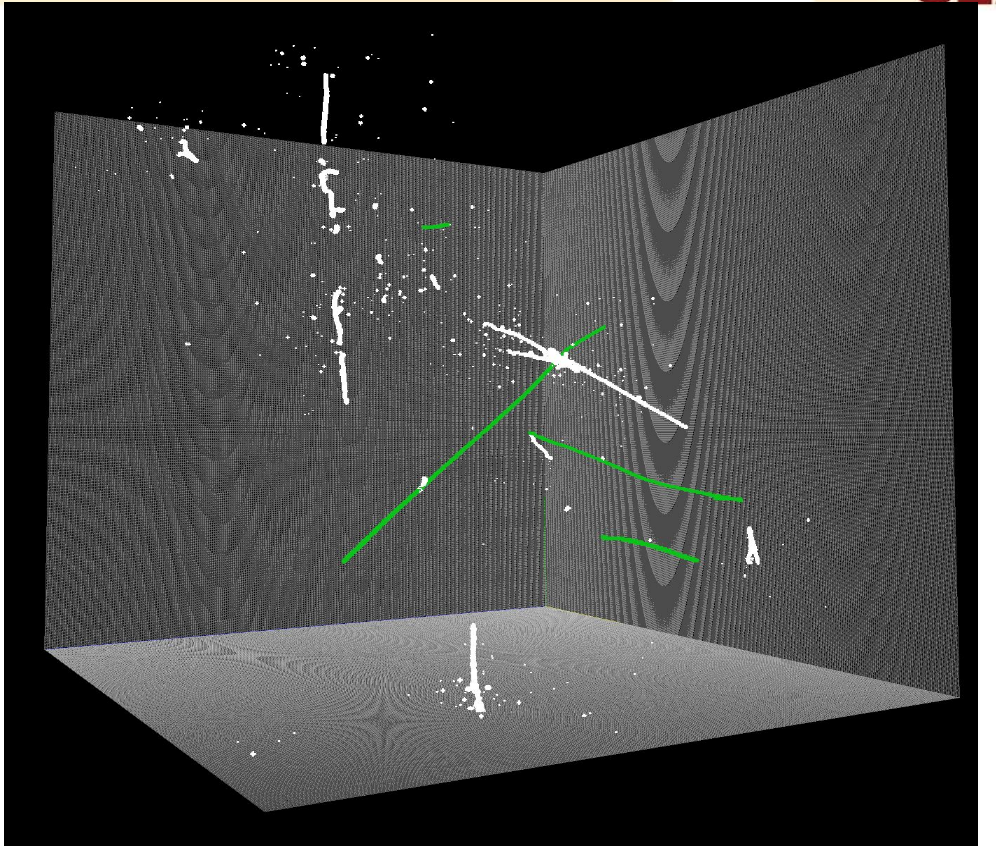
In more details: 2 new operations

- Sparse convolutions (**SC**)
  - Discards contribution of non-active input sites
  - Output site active if at least one input site is active
- Sparse submanifold convolutions (**SSC**)
  - Output size = Input size
  - Output site active iff center of receptive field active
  - Only compute features for active output sites



Scalable CNN for Sparse Particle Imaging Data

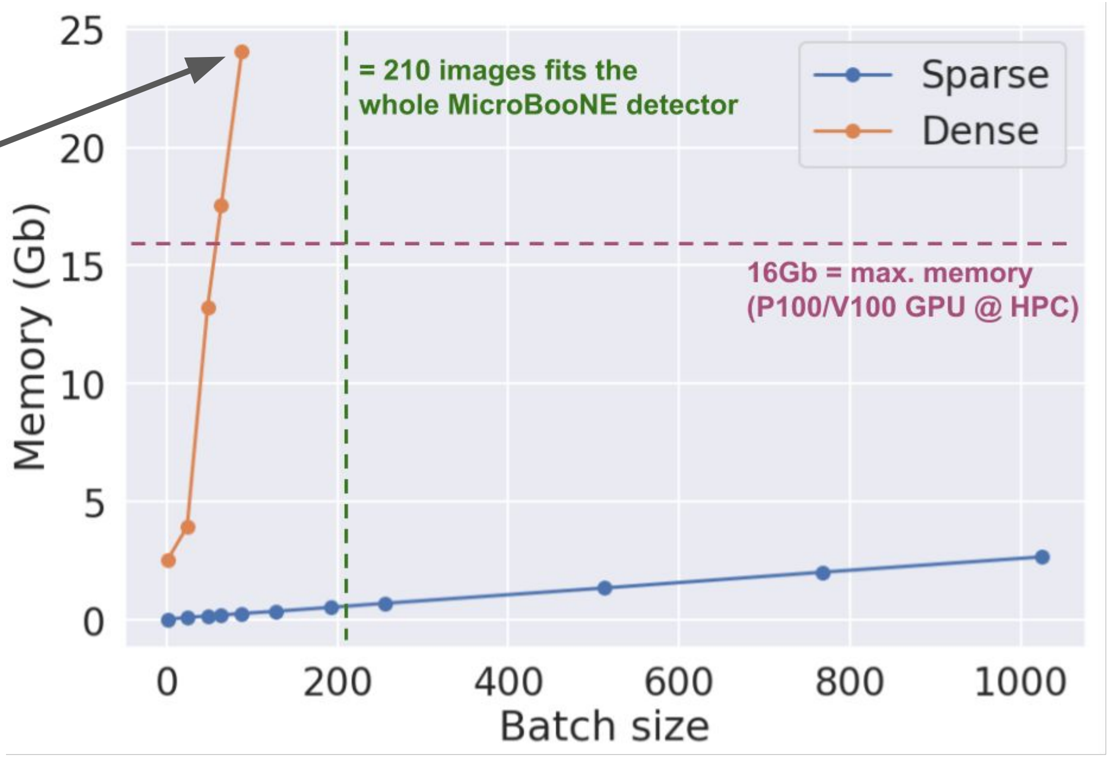
Our data is locally much more dense than ShapeNet 3D dataset



Scalable CNN for Sparse Particle Imaging Data

Sparse U-ResNet fits more data in GPU + good scalability

@batch size 88
sparse uses
93x less memory
than dense and
computation is
3x faster



Scalable CNN for Sparse Particle Imaging Data

SLAC

Sparse Sub-manifold Convolutional NN

- **Public LArTPC simulation**
  - Particle tracking (Geant4) + diffusion, no noise, true energy

Type	Proton	Mu/Pi	Shower	Delta	Michel
Acc.	0.99	0.98	0.99	0.97	0.96

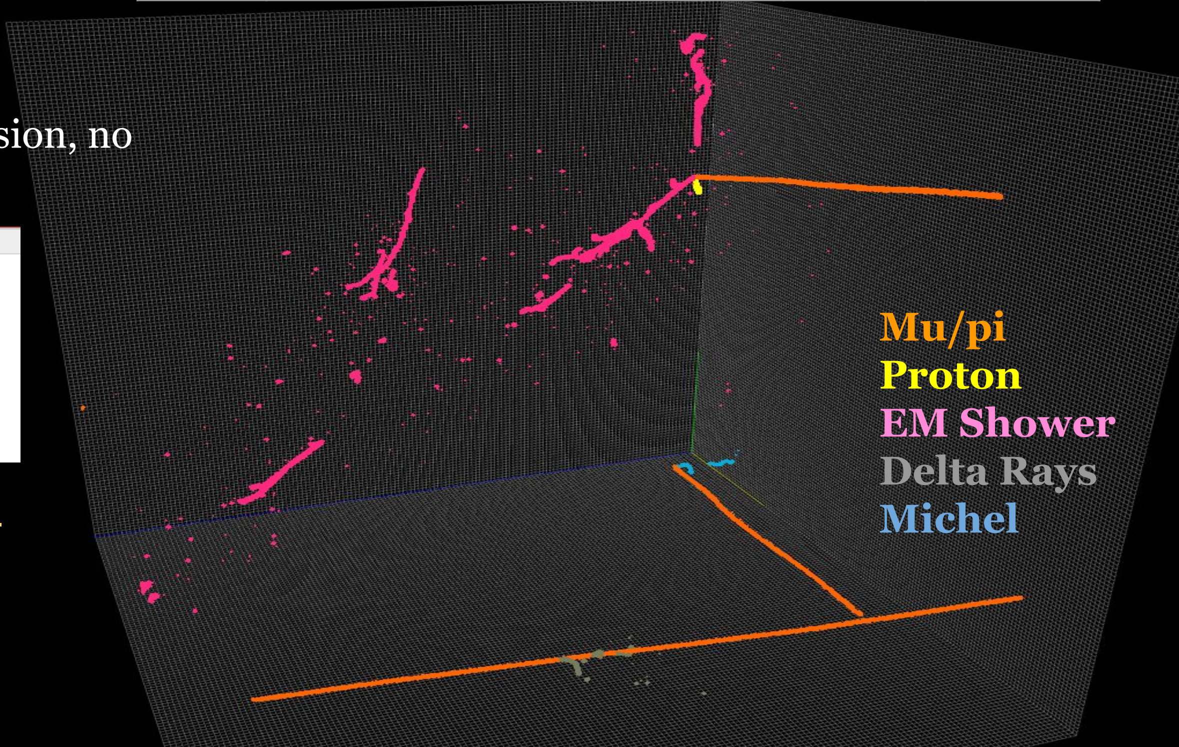
Computer Science > Computer Vision and Pattern Recognition

Scalable Deep Convolutional Neural Networks for Sparse, Locally Dense Liquid Argon Time Projection Chamber Data

Laura Dominé, Kazuhiro Terao

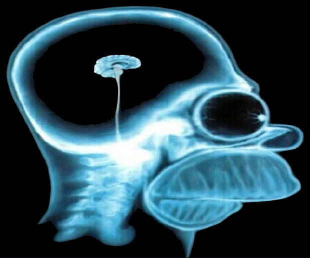
(Submitted on 13 Mar 2019 (v1), last revised 15 Mar 2019 (this version, v2))

Deep convolutional neural networks (CNNs) show strong promise for analyzing scientific data in many domains including particle imaging detectors such as a liquid argon time projection chamber (LArTPC). Yet the high sparsity of LArTPC data challenges traditional CNNs which were designed for dense data such as photographs. A naive application of CNNs on LArTPC data results in inefficient computations and a poor scalability to large LArTPC detectors such as the Short Baseline

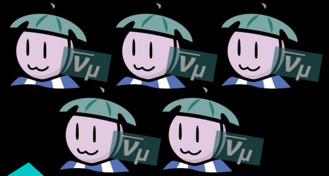


arXiv:1903.05663 presented @ ACAT 2019

- Memory reduction ~ 1/360
- Compute time ~ 1/30
- Handles large future detectors



ML-Based LArTPC Data Reconstruction

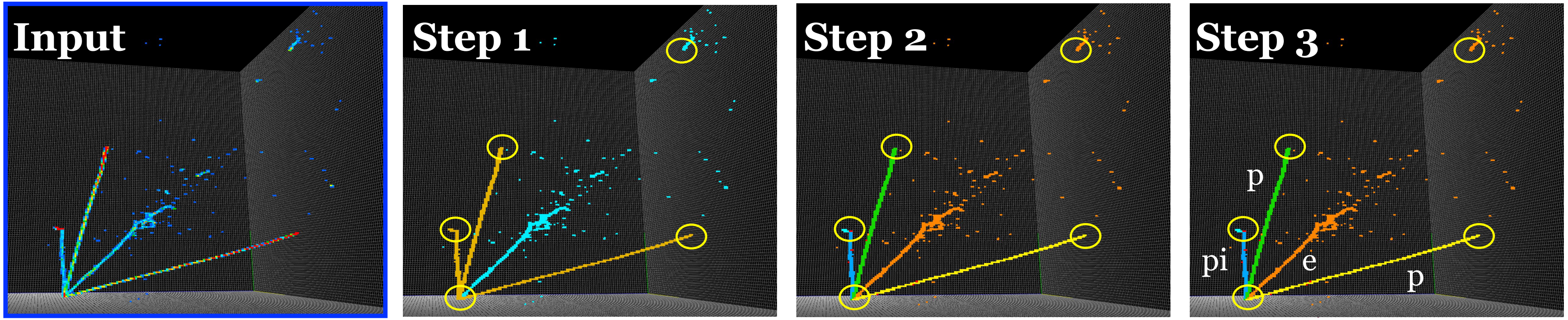


Data Reconstruction Chain

Extraction of hierarchical 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

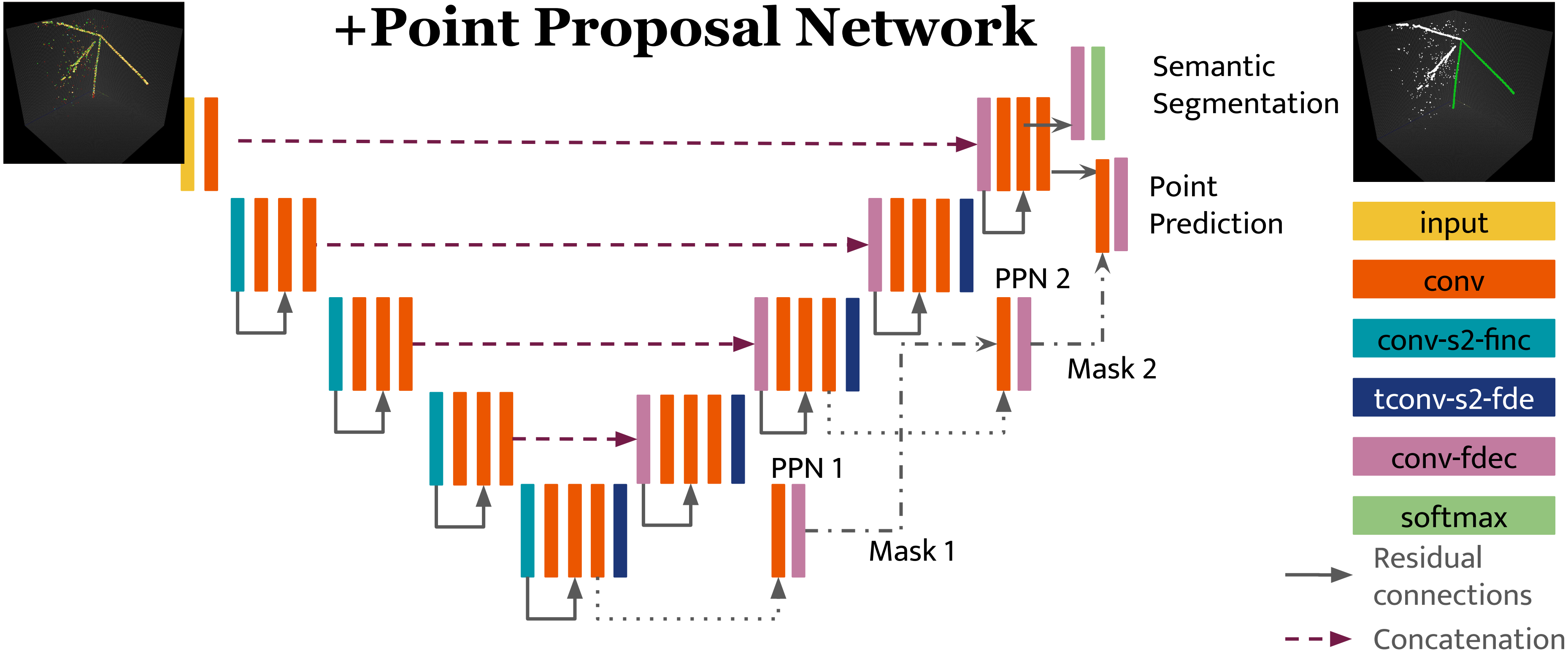
Make it for Hi-resolution 3D image data



ML-based Neutrino Data Reconstruction Chain

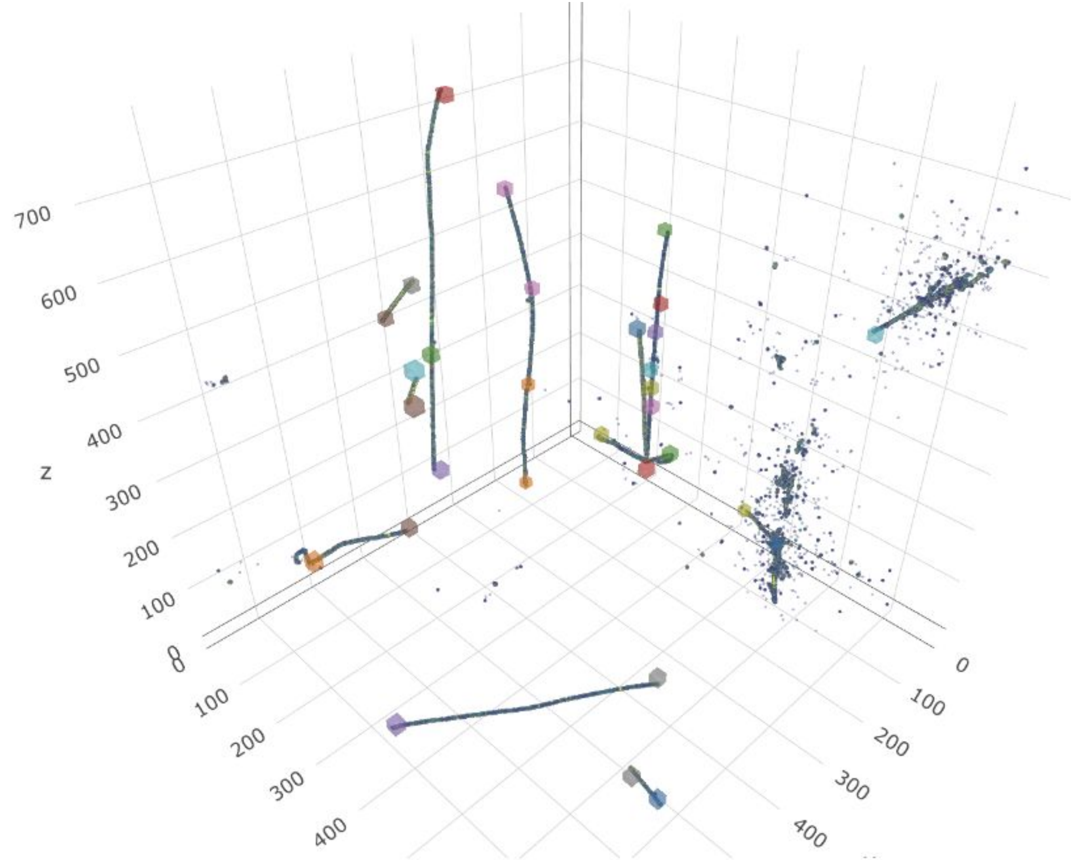
Stage 1: Hi-Res + Abstract Feature Extraction

Architecture: **U-Net + Residual Connections +Point Proposal Network**



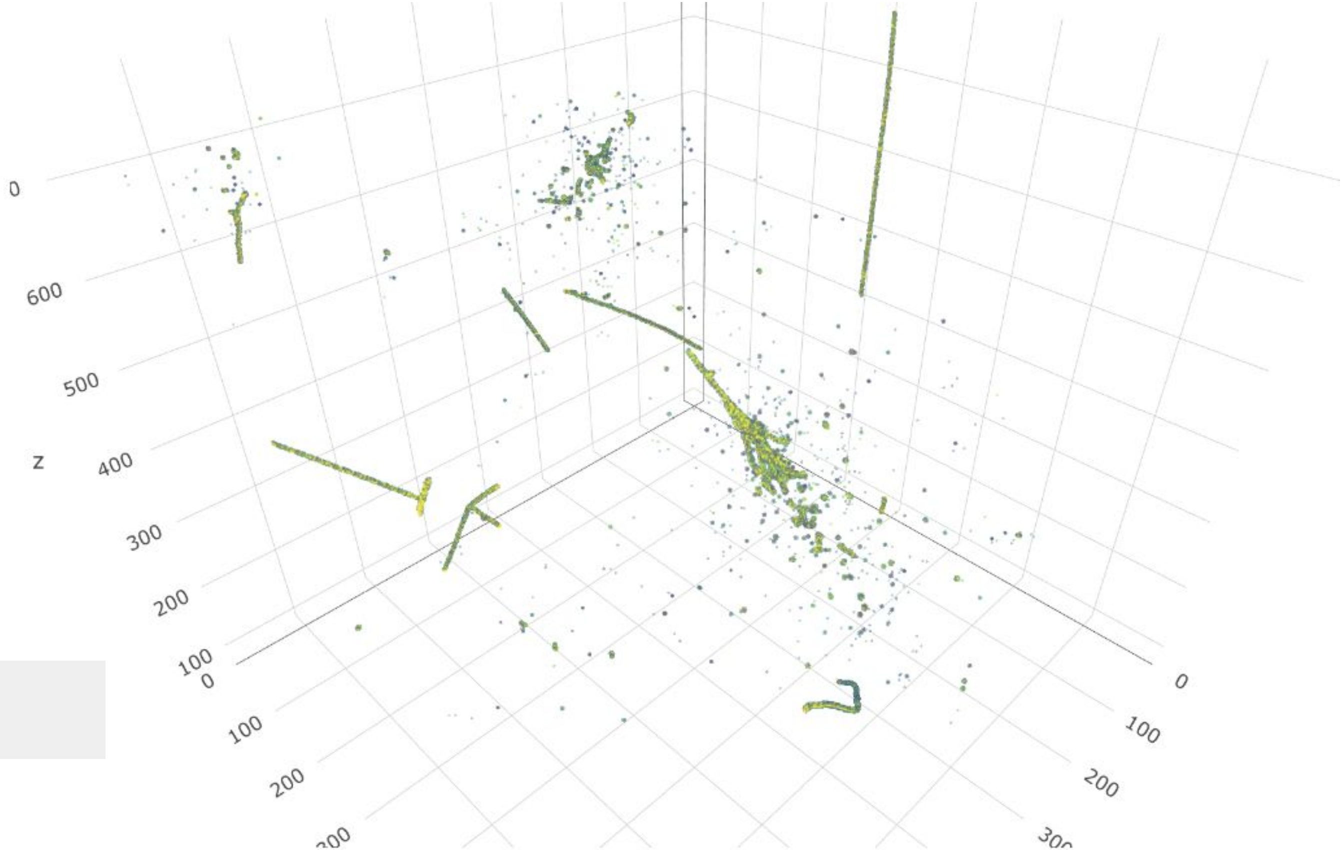
ML-based Neutrino Data Reconstruction Chain

Stage 1: Hi-Res + Abstract Feature Extraction



Deep Proposal

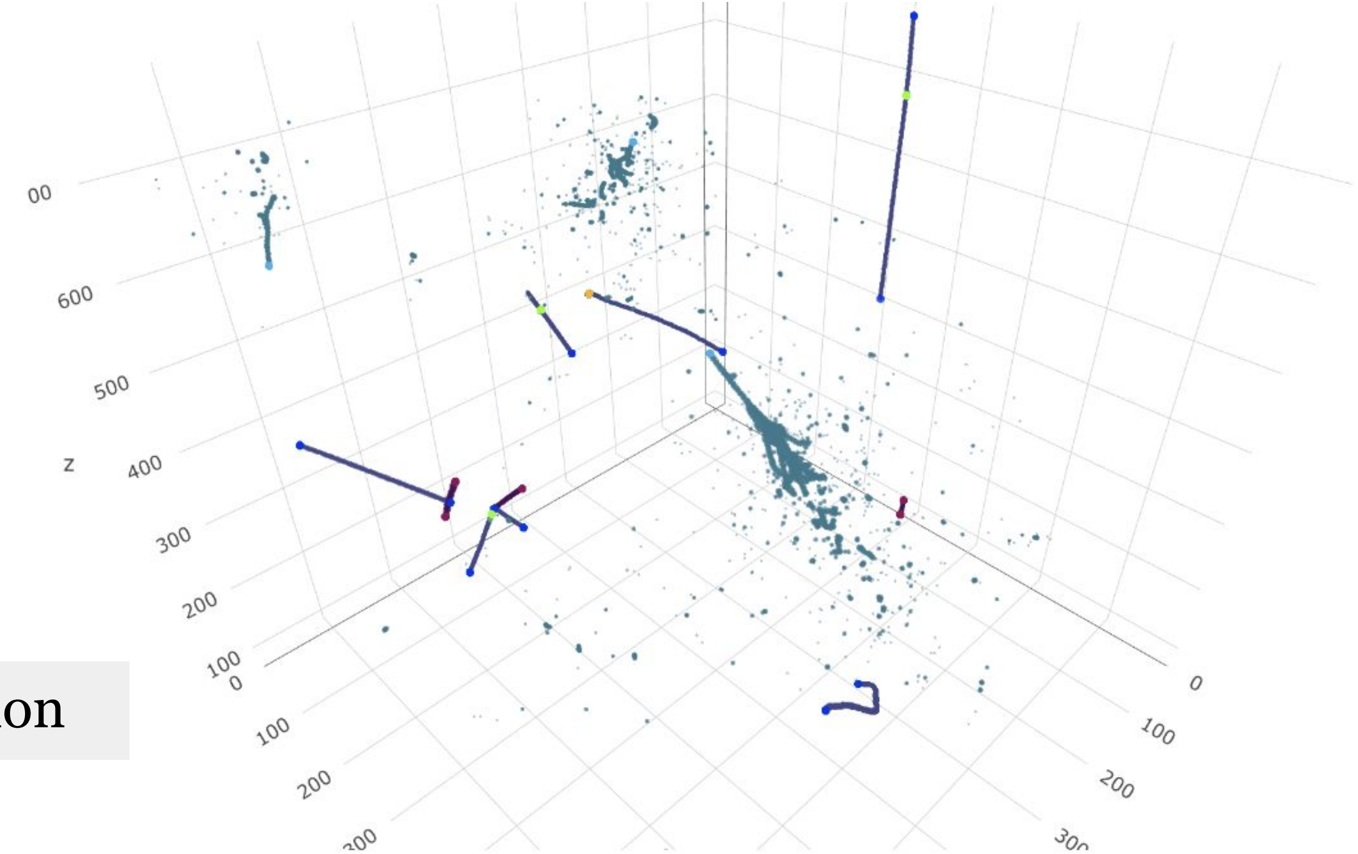
ML-based Neutrino Data Reconstruction Chain
Stage 1: Hi-Res + Abstract Feature Extraction



Input

ML-based Neutrino Data Reconstruction Chain

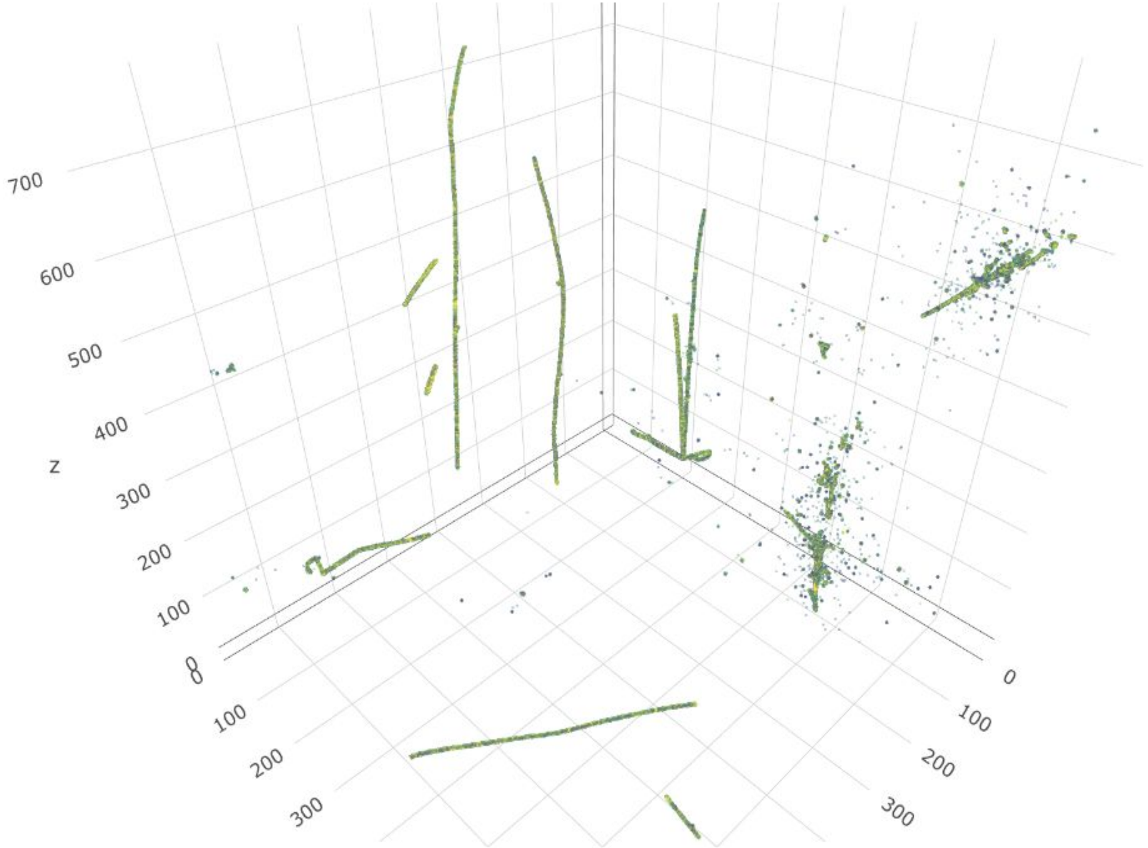
Stage 1: Hi-Res + Abstract Feature Extraction



Prediction

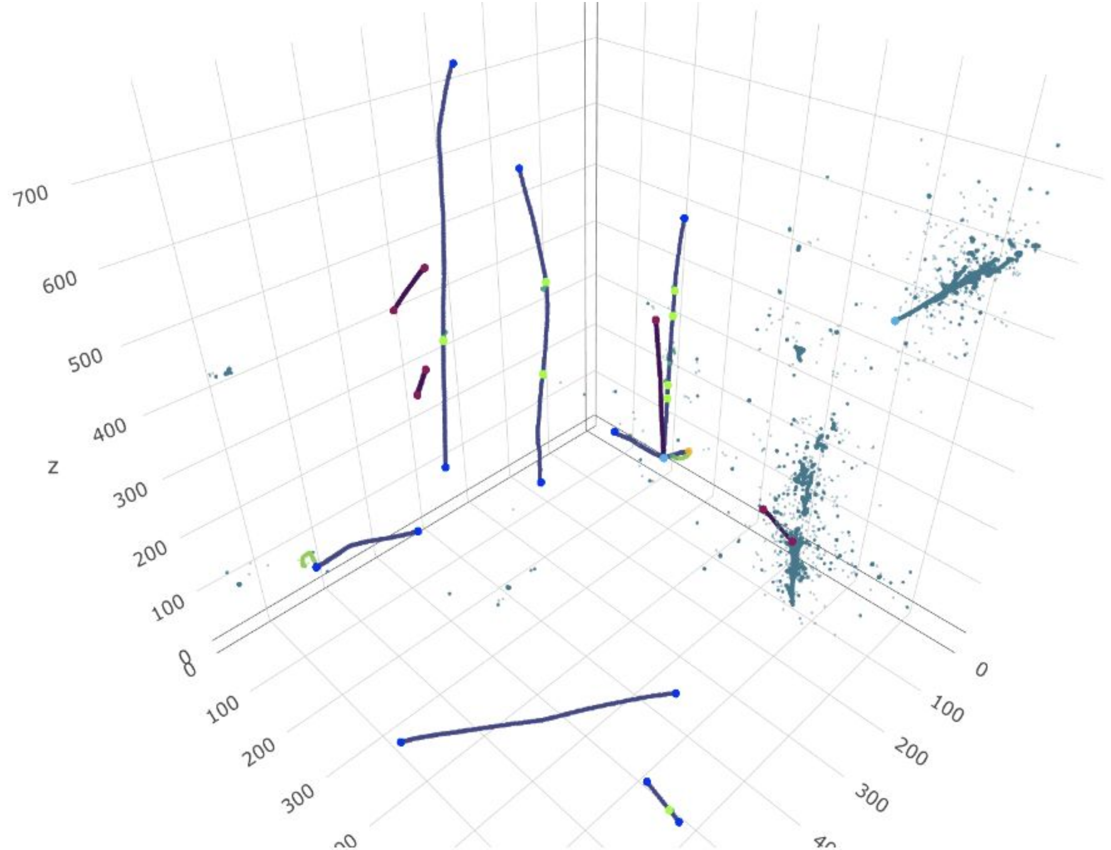
ML-based Neutrino Data Reconstruction Chain

Stage 1: Hi-Res + Abstract Feature Extraction



Input

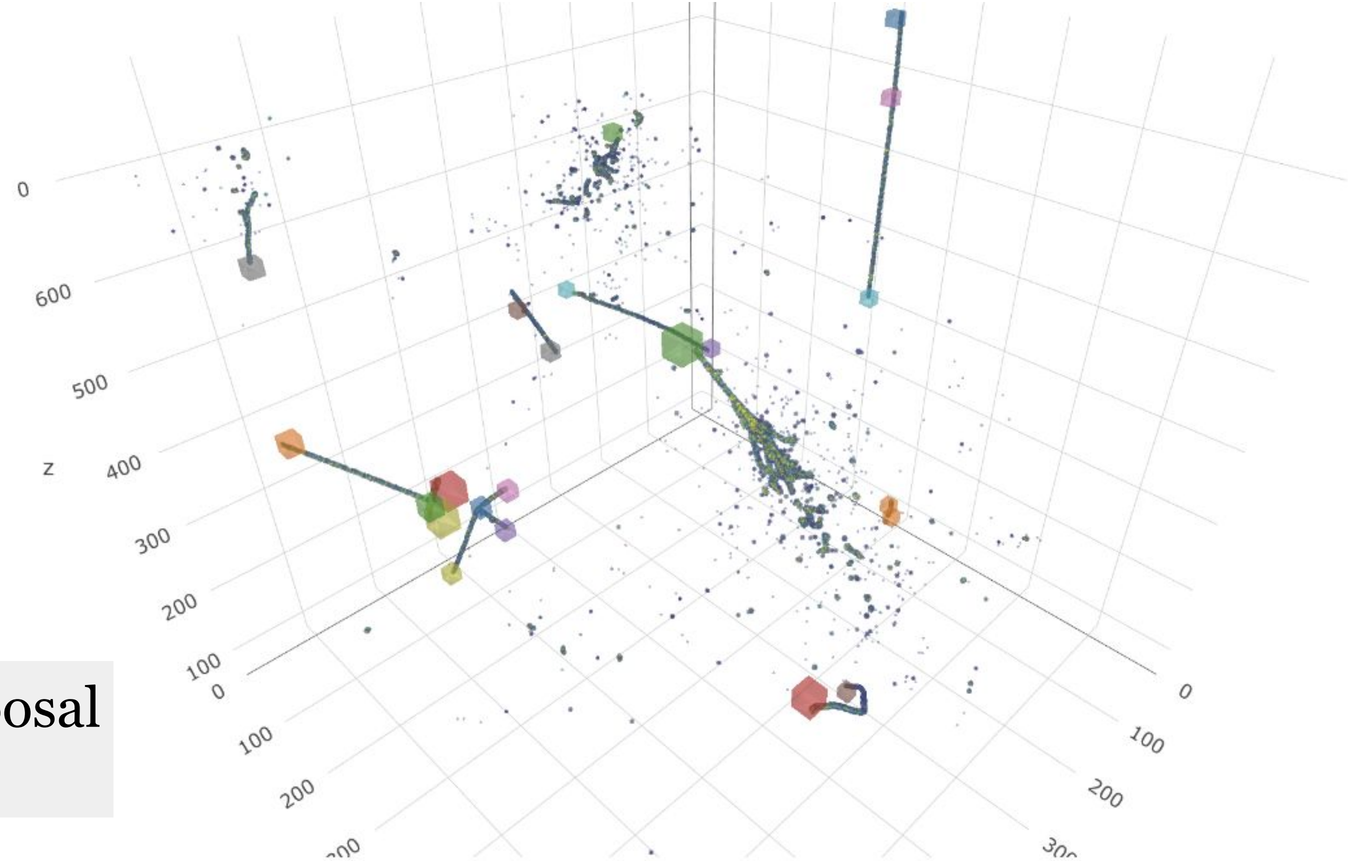
ML-based Neutrino Data Reconstruction Chain
Stage 1: Hi-Res + Abstract Feature Extraction



Prediction

ML-based Neutrino Data Reconstruction Chain

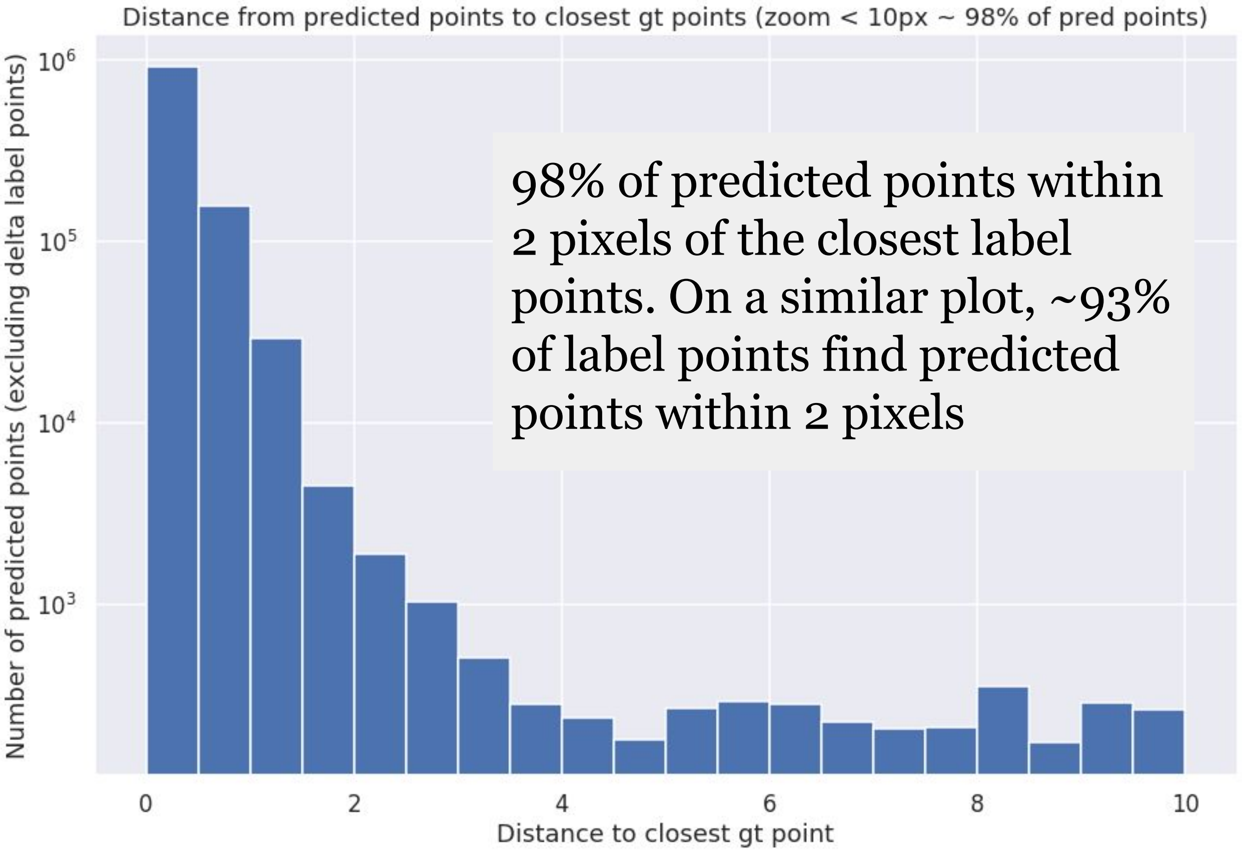
Stage 1: Hi-Res + Abstract Feature Extraction



Deep Proposal

ML-based Neutrino Data Reconstruction Chain

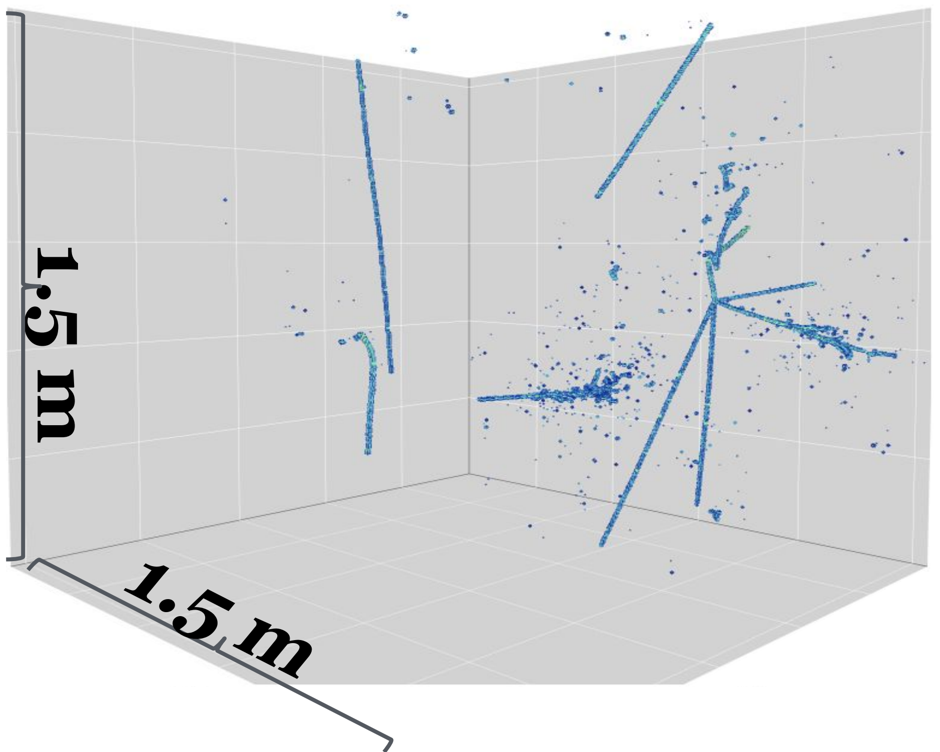
Stage 1: Hi-Res + Abstract Feature Extraction



98% of predicted points within 2 pixels of the closest label points. On a similar plot, ~93% of label points find predicted points within 2 pixels

ML-based Neutrino Data Reconstruction Chain
Stage 2: Particle & Interaction Clustering

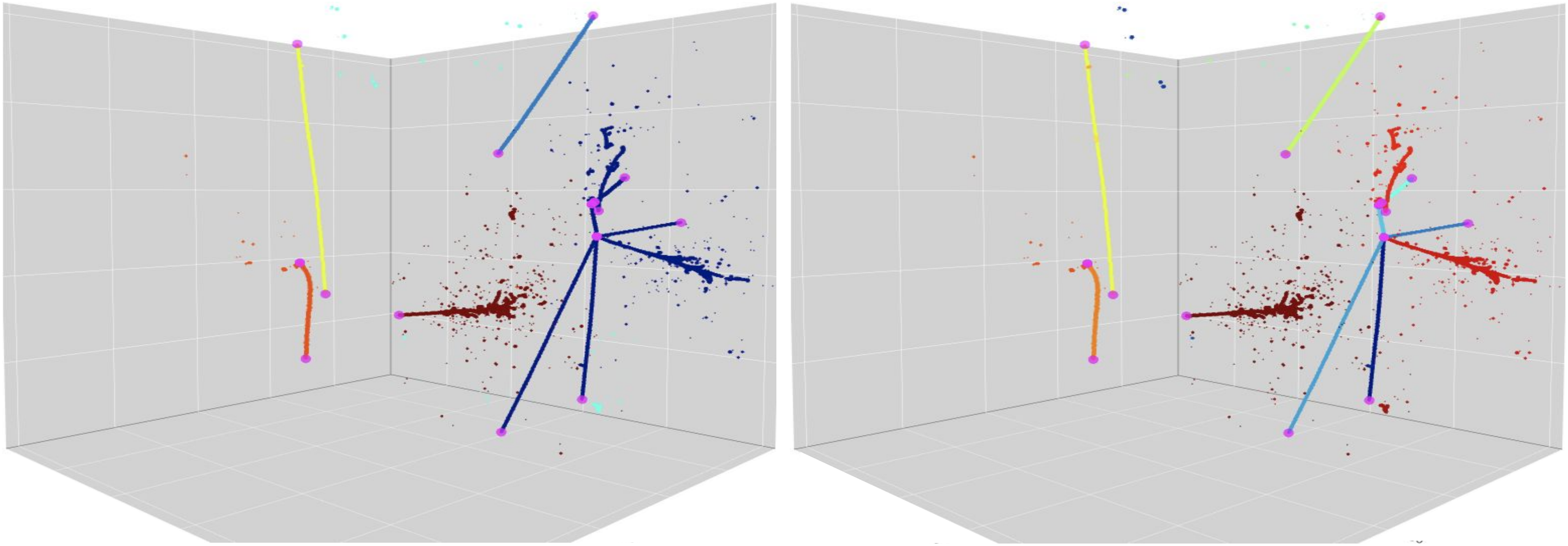
Goal: group pixels into interesting unit of instance



ML-based Neutrino Data Reconstruction Chain

Stage 2: Particle & Interaction Clustering

Goal: group pixels into interesting unit of instance

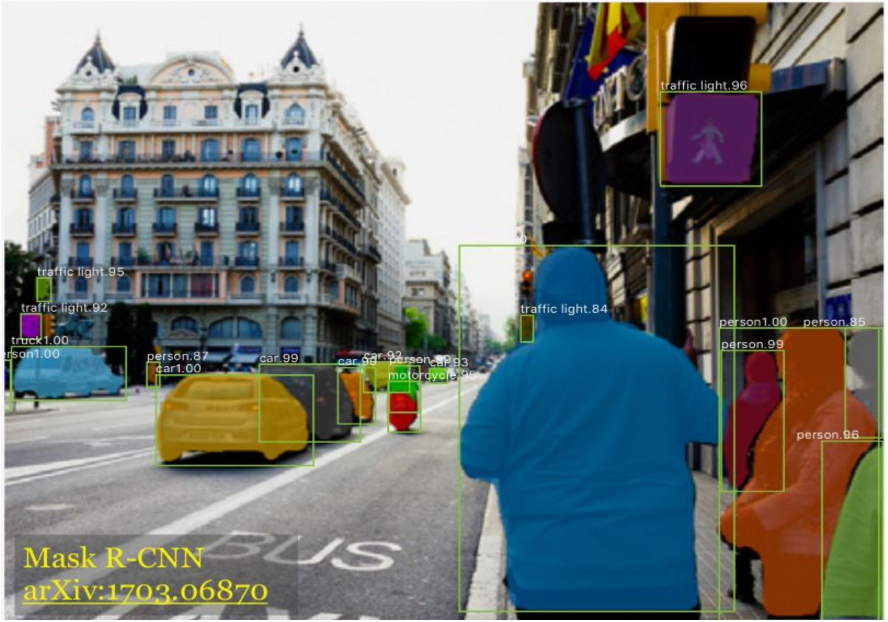


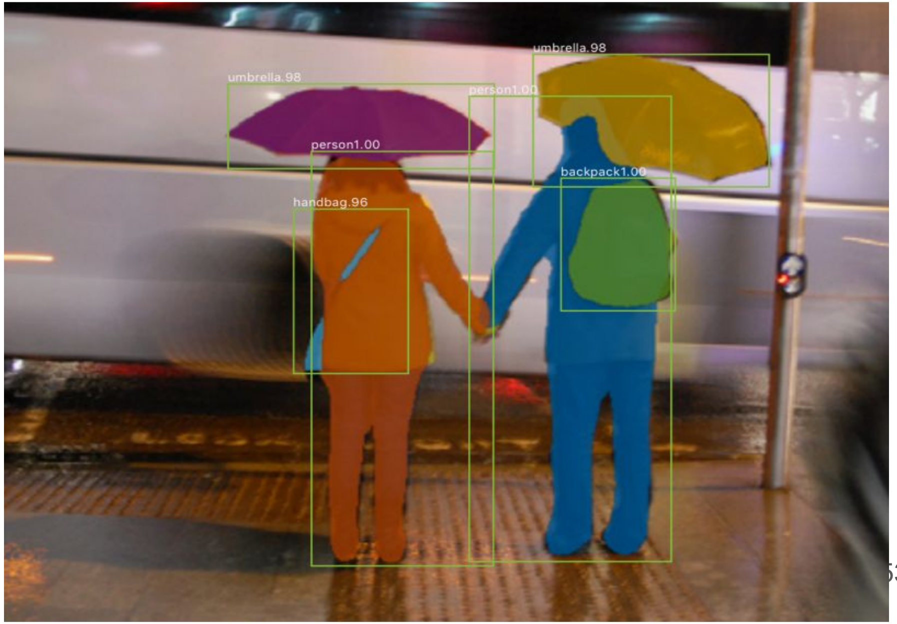
Interaction **Particle**

Stage 2: Particle & Interaction Clustering

Jargon: **Instance (-aware) Semantic Segmentation**

- **Mask R-CNN** ... most popular in industries
  - Object detection + 0/1 instance pixel masking inside each box

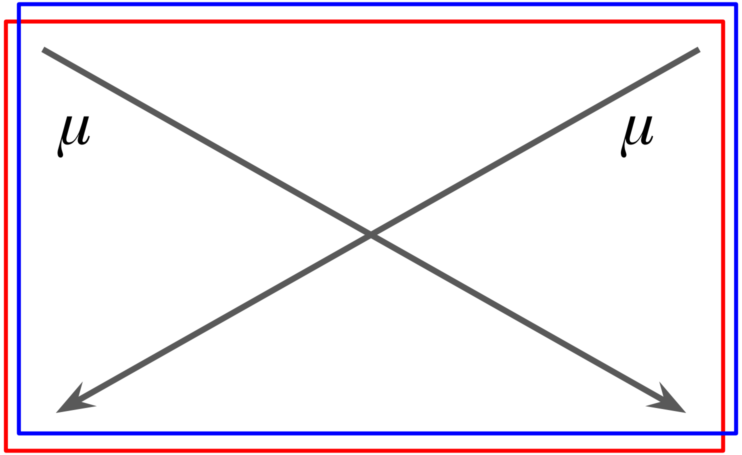




Stage 2: Particle & Interaction Clustering

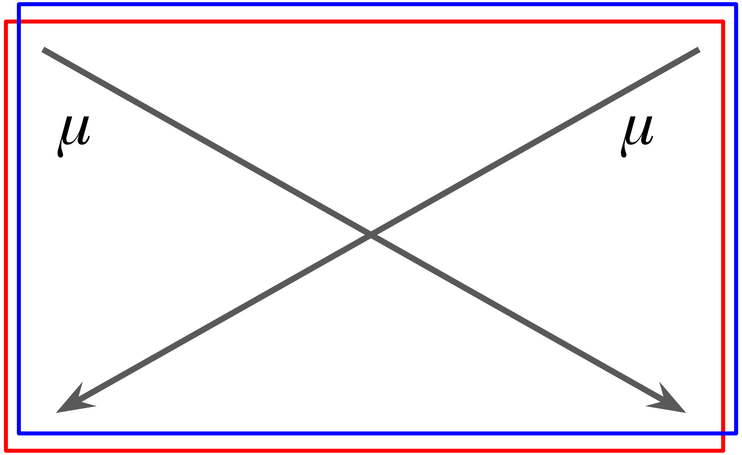
Jargon: **Instance (-aware) Semantic Segmentation**

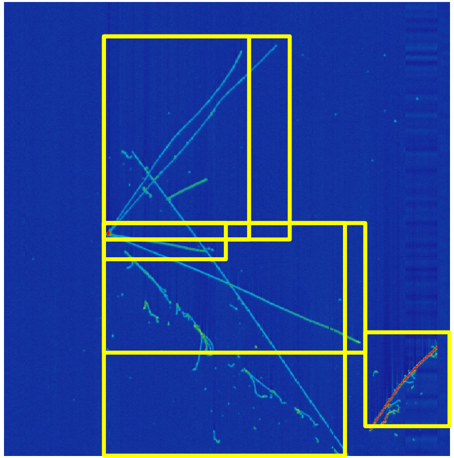
- **Mask R-CNN** … most popular in industries
  - Object detection + 0/1 instance pixel masking in each bounding box (BB)
  - Based on Faster R-CNN (+ ROI-Align + instance masking layers)
  - **Issue**: instance distinction is strongly based on unique BB position/size



Jargon: **Instance (-aware) Semantic Segmentation**

- **Mask R-CNN** … most popular in industries
  - Object detection + 0/1 instance pixel masking in each bounding box (BB)
  - Based on Faster R-CNN (+ ROI-Align + instance masking layers)
  - **Issue**: instance distinction is strongly based on unique BB position/size





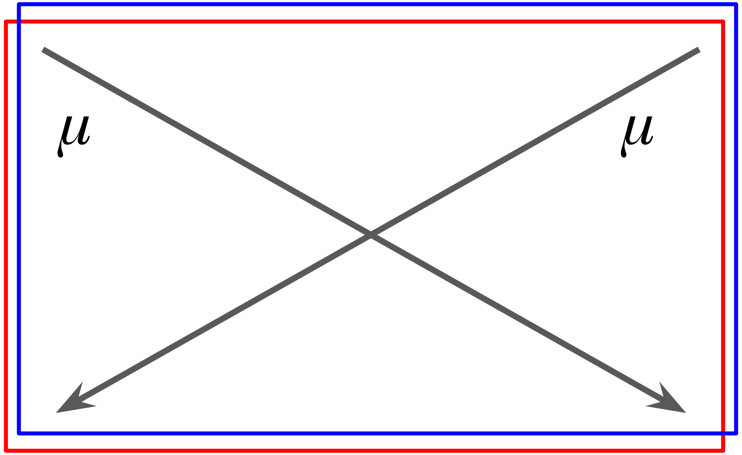
Occlusion issue

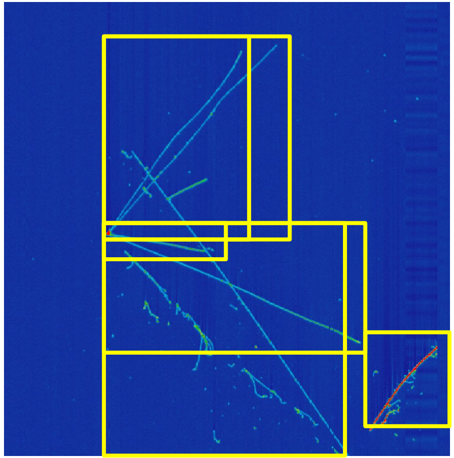
The overlap rate of particles is very high especially for our signal (neutrinos) with an event vertex.

Stage 2: Particle & Interaction Clustering

Jargon: **Instance (-aware) Semantic Segmentation**

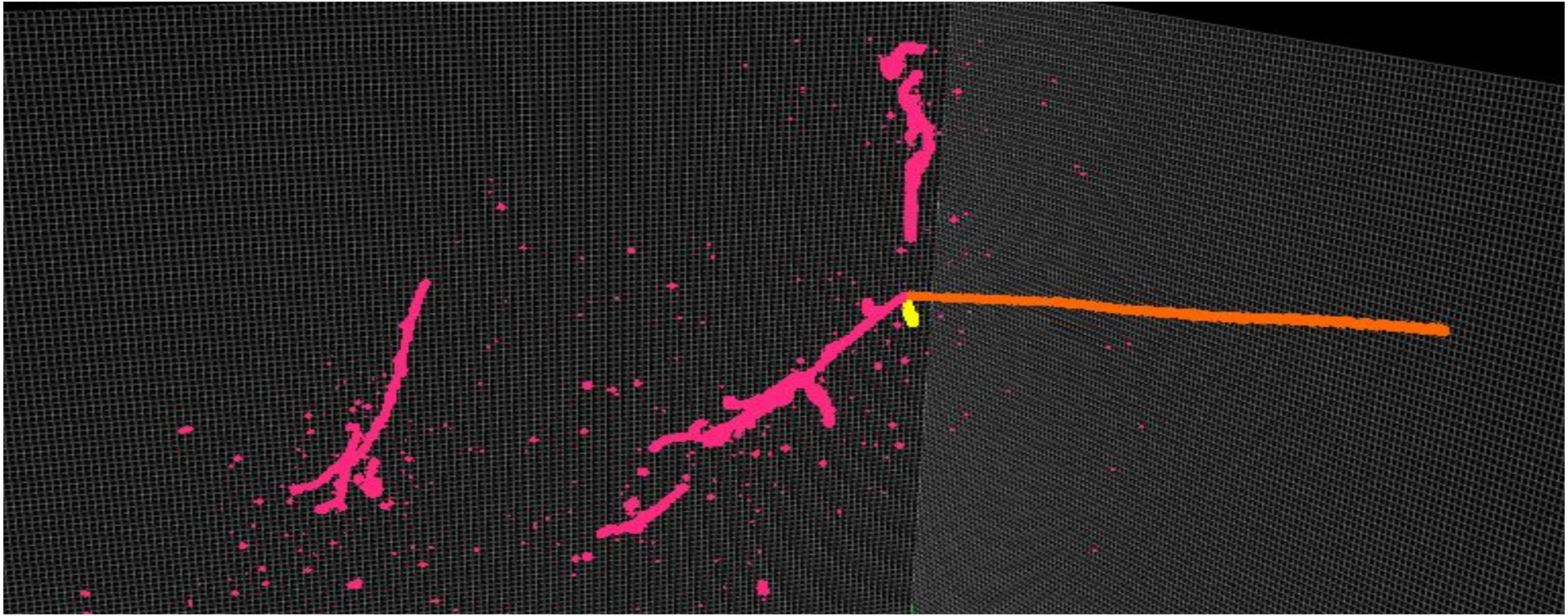
- **Mask R-CNN** … most popular in industries
  - Object detection + 0/1 instance pixel masking in each bounding box (BB)
  - Based on Faster R-CNN (+ ROI-Align + instance masking layers)
  - **Issue**: instance distinction is strongly based on unique BB position/size





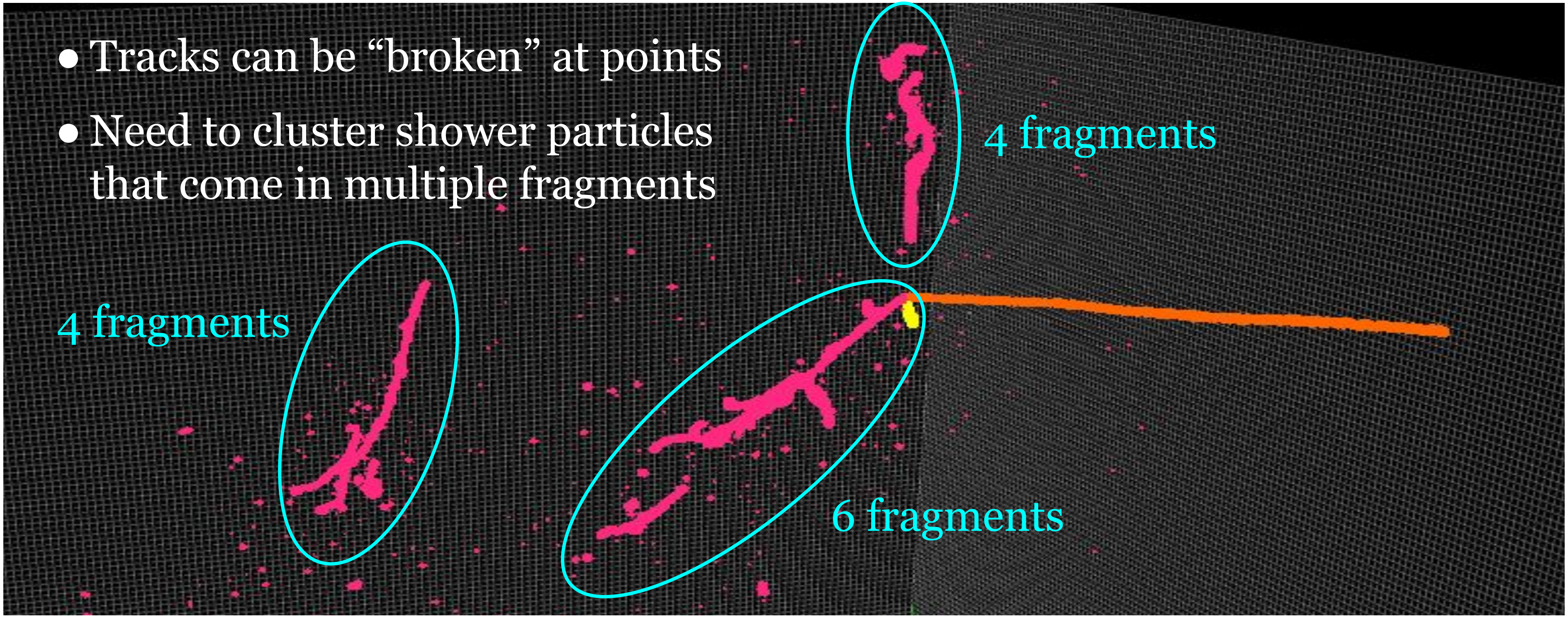
Stage 2: Particle & Interaction Clustering

Alternative 1: cluster segmented fragments



Stage 2: Particle & Interaction Clustering

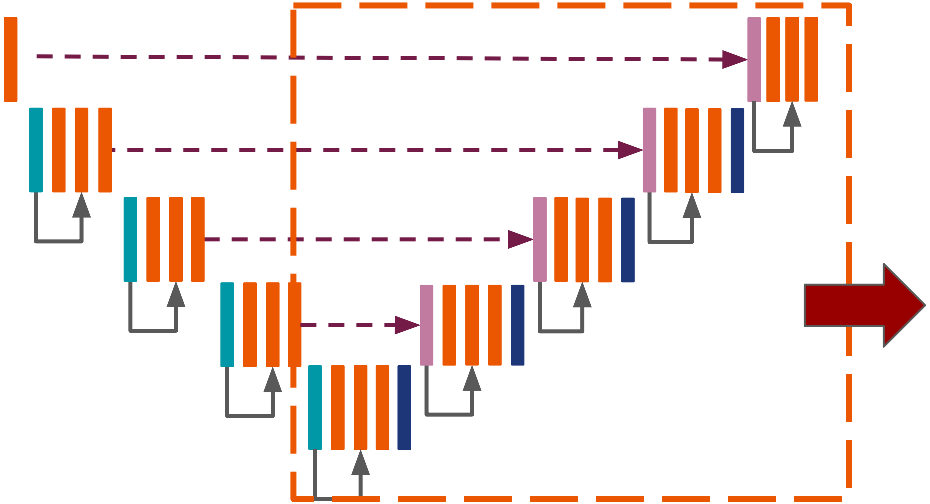
Alternative 1: cluster segmented fragments



Stage 2: Particle & Interaction Clustering

Alternative 1: cluster segmented fragments

- **Graph Neural Networks**
    - Define cluster fragments (nodes) by DBSCAN per segmentation mask
    - Construct node features (re-use multi-scale features already extracted)

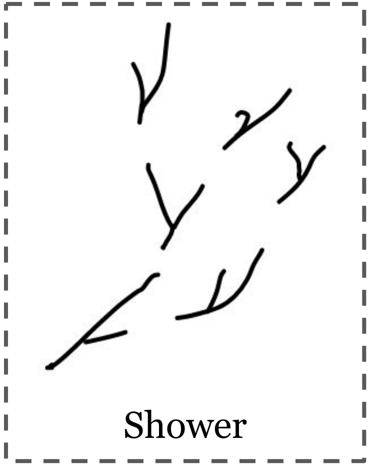


ala Feature Pyramid
Per fragment, apply mask at each scale + pooling to define the same node tensor shape

ML-based Neutrino Data Reconstruction Chain
Stage 2: Particle & Interaction Clustering

Alternative 1: cluster segmented fragments

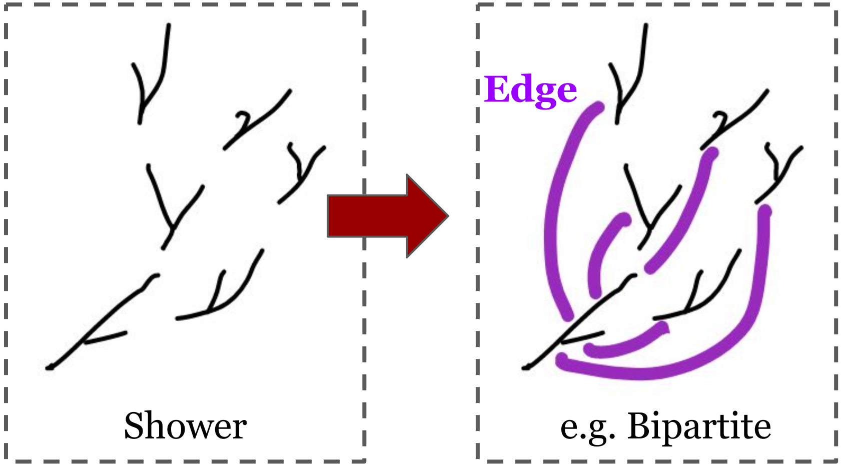
- **Graph Neural Networks**
  - Define cluster fragments (nodes) by DBSCAN per segmentation mask
  - Construct node features (re-use multi-scale features already extracted)
  - Define possible connections among fragments (edges)

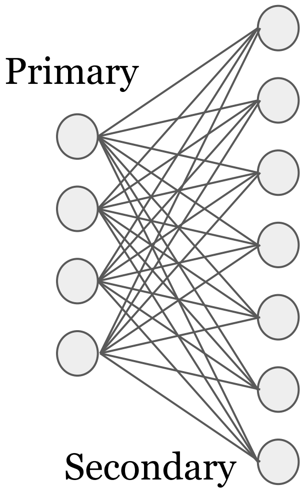


ML-based Neutrino Data Reconstruction Chain
Stage 2: Particle & Interaction Clustering

Alternative 1: cluster segmented fragments

- **Graph Neural Networks**
    - Define cluster fragments (nodes) by DBSCAN per segmentation mask
    - Construct node features (re-use multi-scale features already extracted)
    - Define possible connections among fragments (edges)





- "Primary" is first shower fragment
- NxM edges is not too large to handle
- Some edges may have weak/difficult

ML-based Neutrino Data Reconstruction Chain
Stage 2: Particle & Interaction Clustering

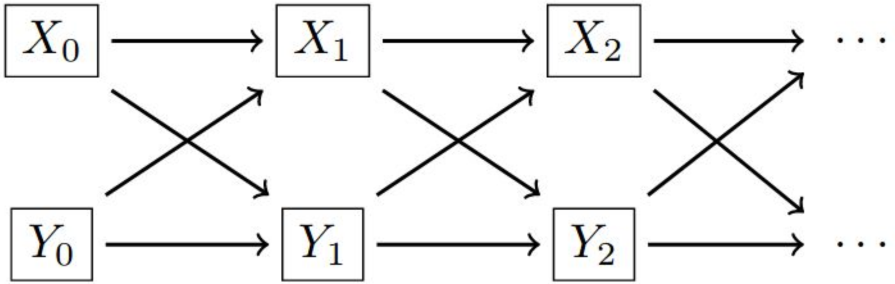
SLAC

Alternative 1: cluster segmented fragments

- **Graph Neural Networks**
  - Define cluster fragments (nodes) by DBSCAN per segmentation mask
  - Construct node features (re-use multi-scale features already extracted)
  - Define possible connections among fragments (edges)

GNN recap (maybe skip?)

- X_k and Y_k are k-th layer node & edge



- Edge feature at (i, j), layer k+1

$$Y_{i,j;k+1} = f(X_{i;k}, X_{j;k}, Y_{i,j;k})$$

- Message from the edge (i, j)

$$M_{i,j;k+1} = f(X_{i;k}, X_{j;k}, Y_{i,j;k})$$

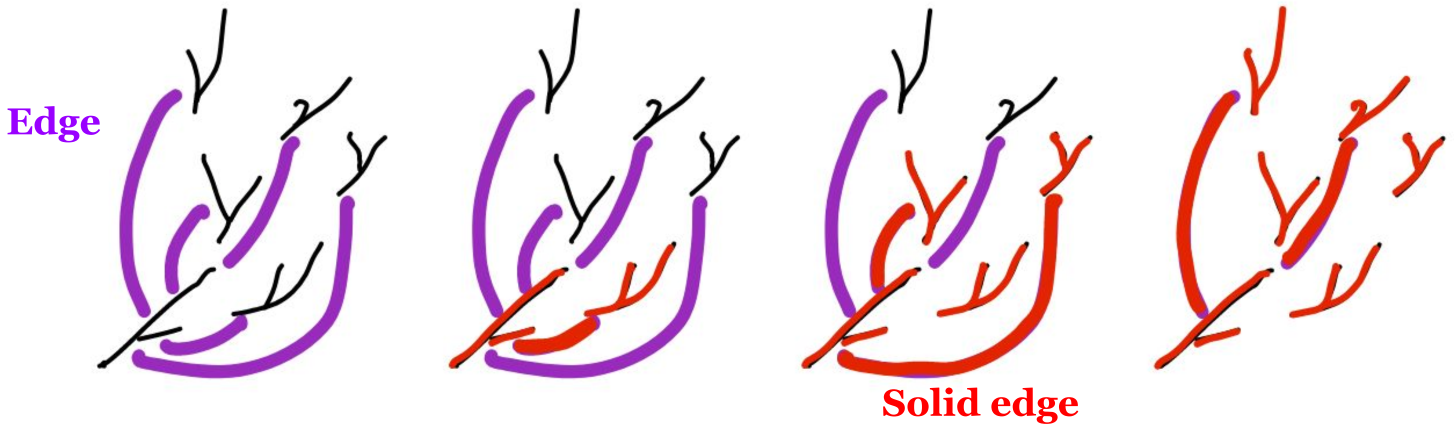
- Node feature at i, layer k+1

$$X_{i;k+1} = \mathop{\mathrm{Op}}_{j \in N(i)} M_{i,j;k+1}$$

Stage 2: Particle & Interaction Clustering

Alternative 1: cluster segmented fragments

- **Dynamic Graph Neural Networks**
  - Define cluster fragments (nodes) by DBSCAN per segmentation mask
  - Construct node features (re-use multi-scale features already extracted)
  - Define possible connections among fragments (edges)



ML-based Neutrino Data Reconstruction Chain
Stage 2: Particle & Interaction Clustering

Alternative 2: transform data into easily clusterable hyperspace

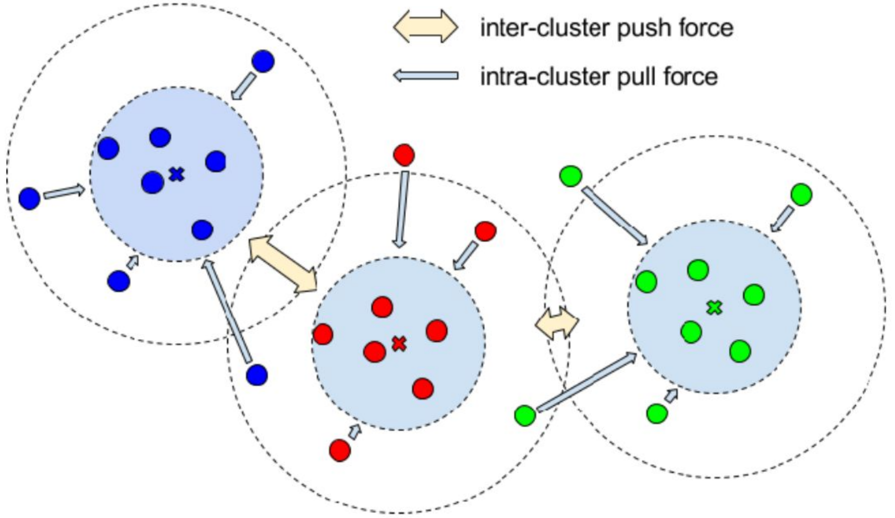
- **GNN or CNN**
  - Interpret node/pixel features from GNN/CNN as hyperspace coordinate

$$L = \alpha L_{var} + \beta L_{dist} + \gamma L_{reg},$$

$$L_{var} = \frac{1}{C}\sum_{c=1}^{C}\frac{1}{N_c}\sum_{i=1}^{N_c}[\max(0, \|\mu_c - x_i\| - \delta_v)]^2$$

$$L_{dist} = \frac{1}{C(C-1)}\sum_{\substack{c_A, c_B = 1 \\ c_A \neq c_B}}^{C}[\max(0, 2\delta_d - \|\mu_{c_A} - \mu_{c_B}\|)]^2$$

$$L_{reg} = \frac{1}{C}\sum_{c=1}^{C}\|\mu_c\|$$



Equation credit: Dae Hyun K. @ Stanford

Image credit: arXiv 1708.02551

ML-based LArTPC Data Reconstruction Chain
Stage 2: Particle & Interaction Clustering

Alternative 2: transform data into easily clusterable hyperspace

- **GNN or CNN**
  - Interpret node/pixel features from GNN/CNN as hyperspace coordinate

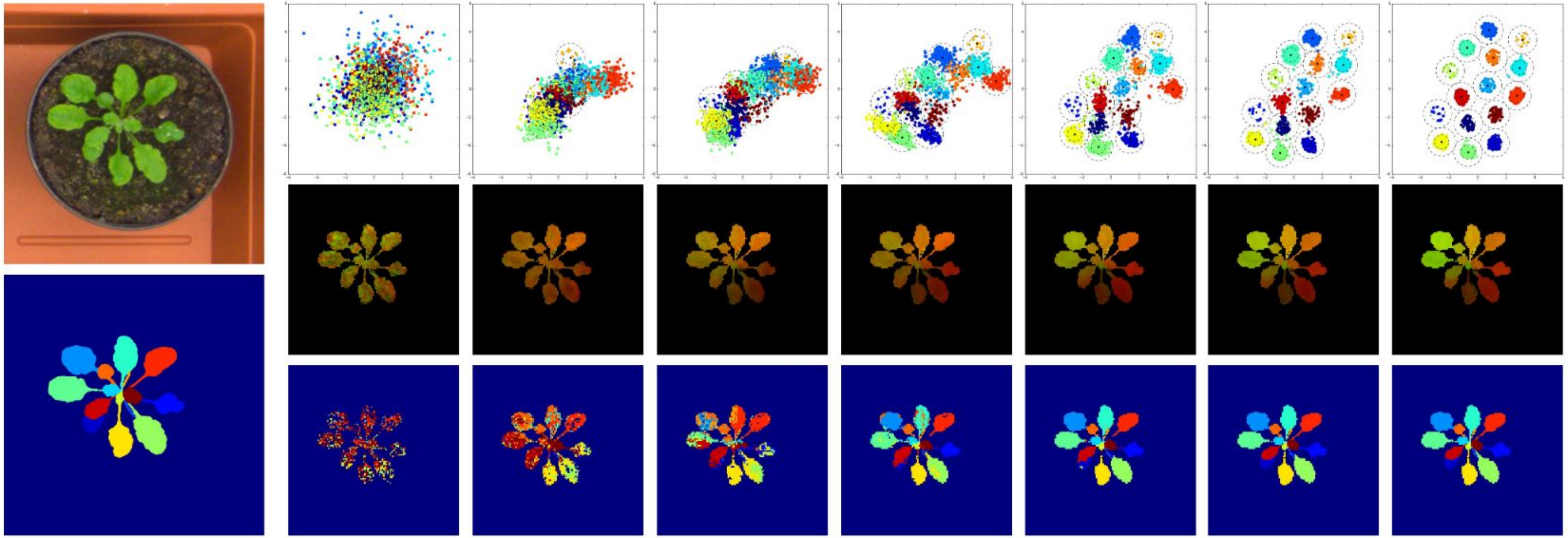


Image credit: arXiv 1708.02551

ML-based LArTPC Data Reconstruction Chain
Stage 2: Particle & Interaction Clustering

Alternative 2: transform data into easily clusterable hyperspace

- **GNN or CNN**
  - Interpret node/pixel features from GNN/CNN as hyperspace coordinate

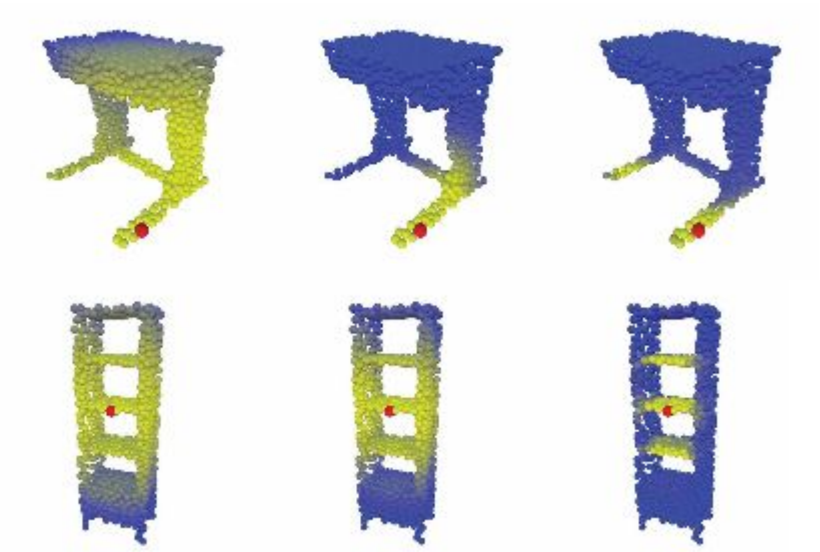
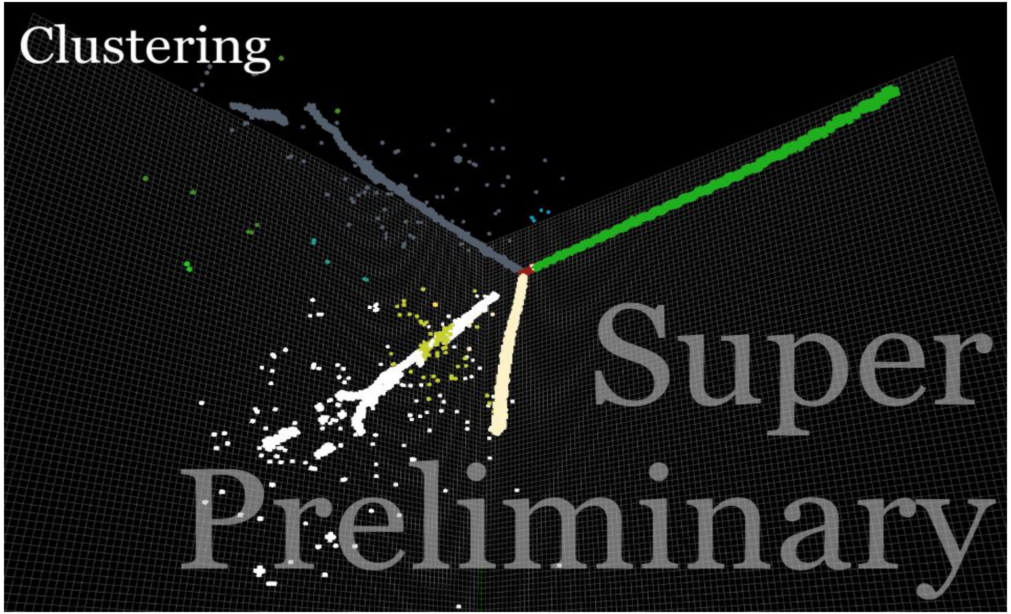


Image credit: arXiv 1801.07829



… wrapping up …

Outline

1. Neutrino Detectors
2. Machine Learning and Computer Vision Applications
3. ML-based Neutrino Data Reconstruction Chain
4. **Summary**

Summary

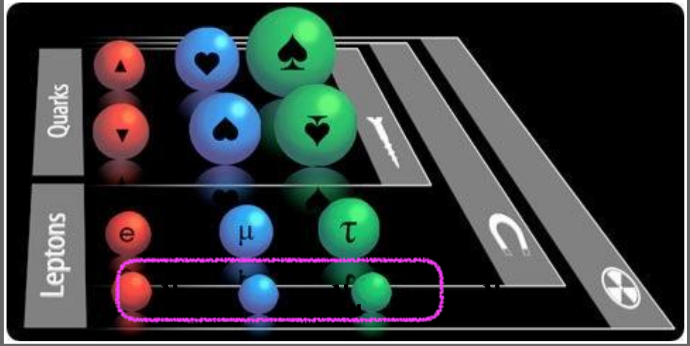
- **Neutrino detector trend: particle imaging**
- **Dedicated image analysis techniques needed**
  - Techniques developed in the field of computer vision, in particular **deep neural networks**, show strong promise
    - Strong synergy = collaboration with scientists beyond HEP
  - **"Data reconstruction" using ML** (my research)
  - Active but not mentioned: data/sim domain adaptation (MINERvA paper), distributed ML on HPCs, etc.
- **I am curious: please tell me about your research :)**

Questions?

Back Up Slides

Why Neutrino Physics? (I)

Standard Model (SM)

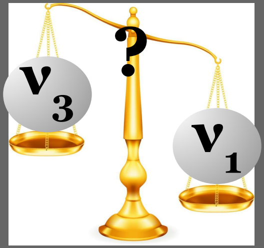


Successful description of how we know particles interact in nature

… but **not so much on neutrinos!**

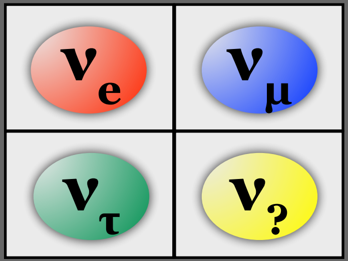
Neutrinos *beyond* SM

With **neutrino oscillations** firmly in place, we know at least there are 3 mass eigenstates. But there is **much more to learn**…



Mass hierarchy
$m_1 > m_3$?

CP violation

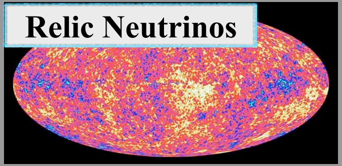


Sterile neutrino?

Why Neutrino Physics? (II)

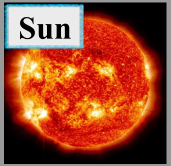
Neutrinos are everywhere

Which makes them **natural probes to the universe and its history**


Relic Neutrinos


SuperNova

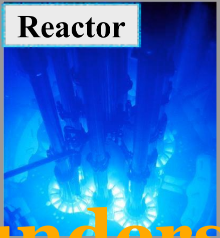

AGN


Sun

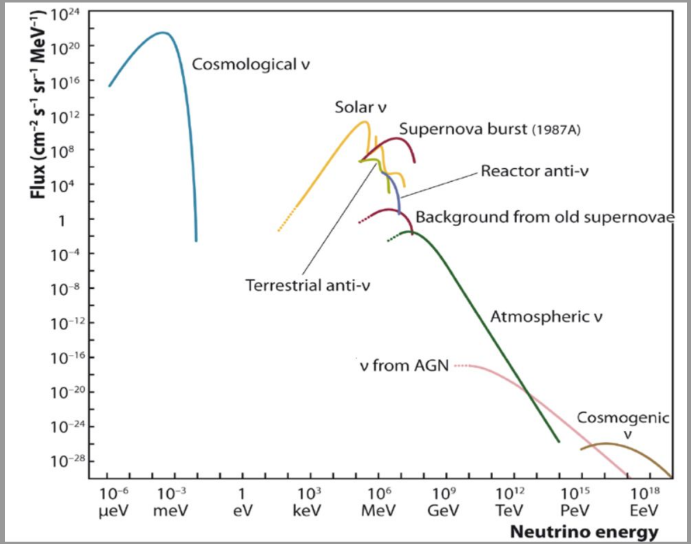

Atmospheric


Earth


Accelerator


Reactor


Good Stuff



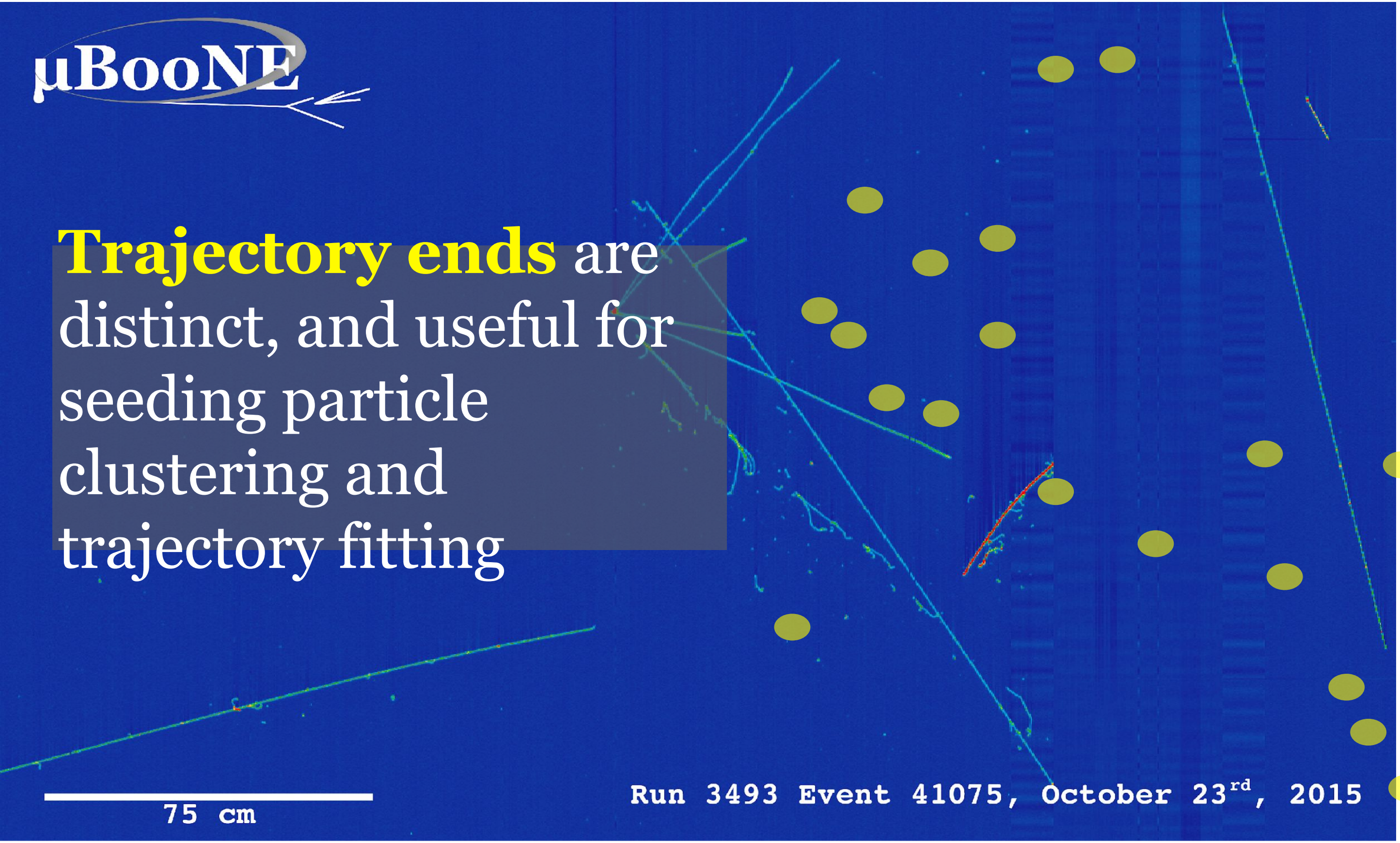
EPJ H37 (2012) 3:515-565

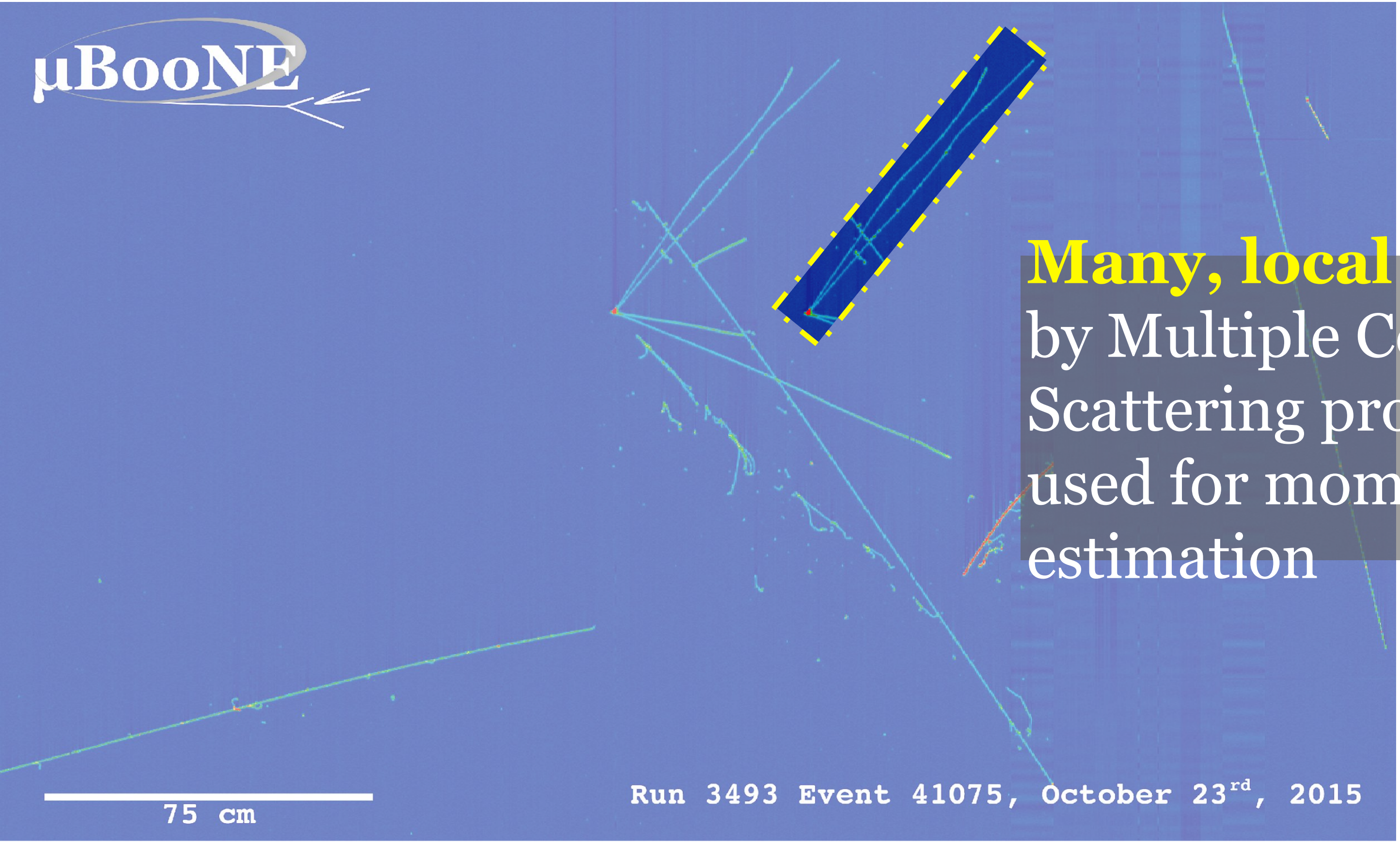
Need to understand more about them!

Oscillation physics has taught us a lot, but still much to learn…

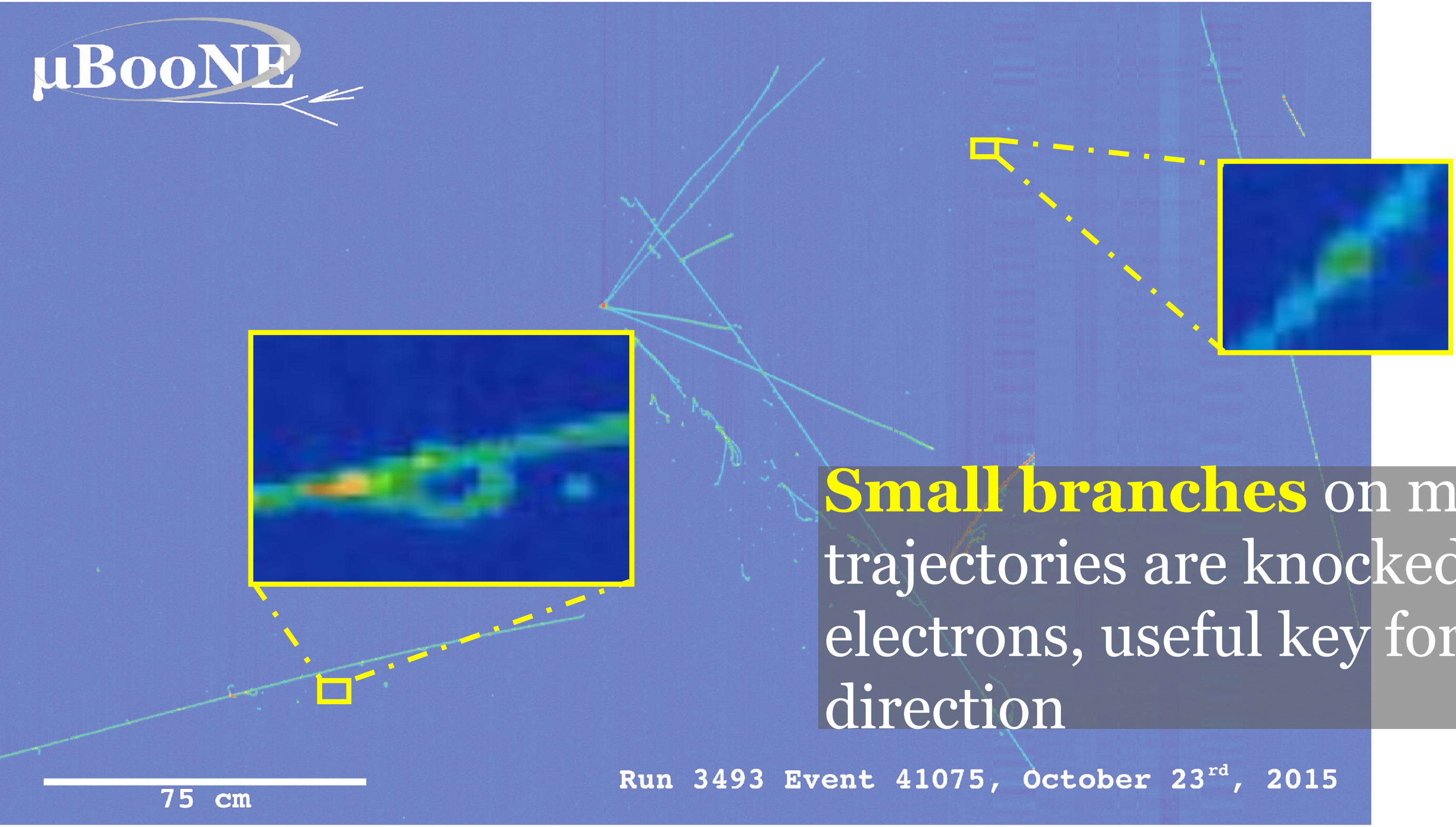
Topological shape difference is a major distinction for “shower” particles

Trajectory ends are distinct, and useful for seeding particle clustering and trajectory fitting



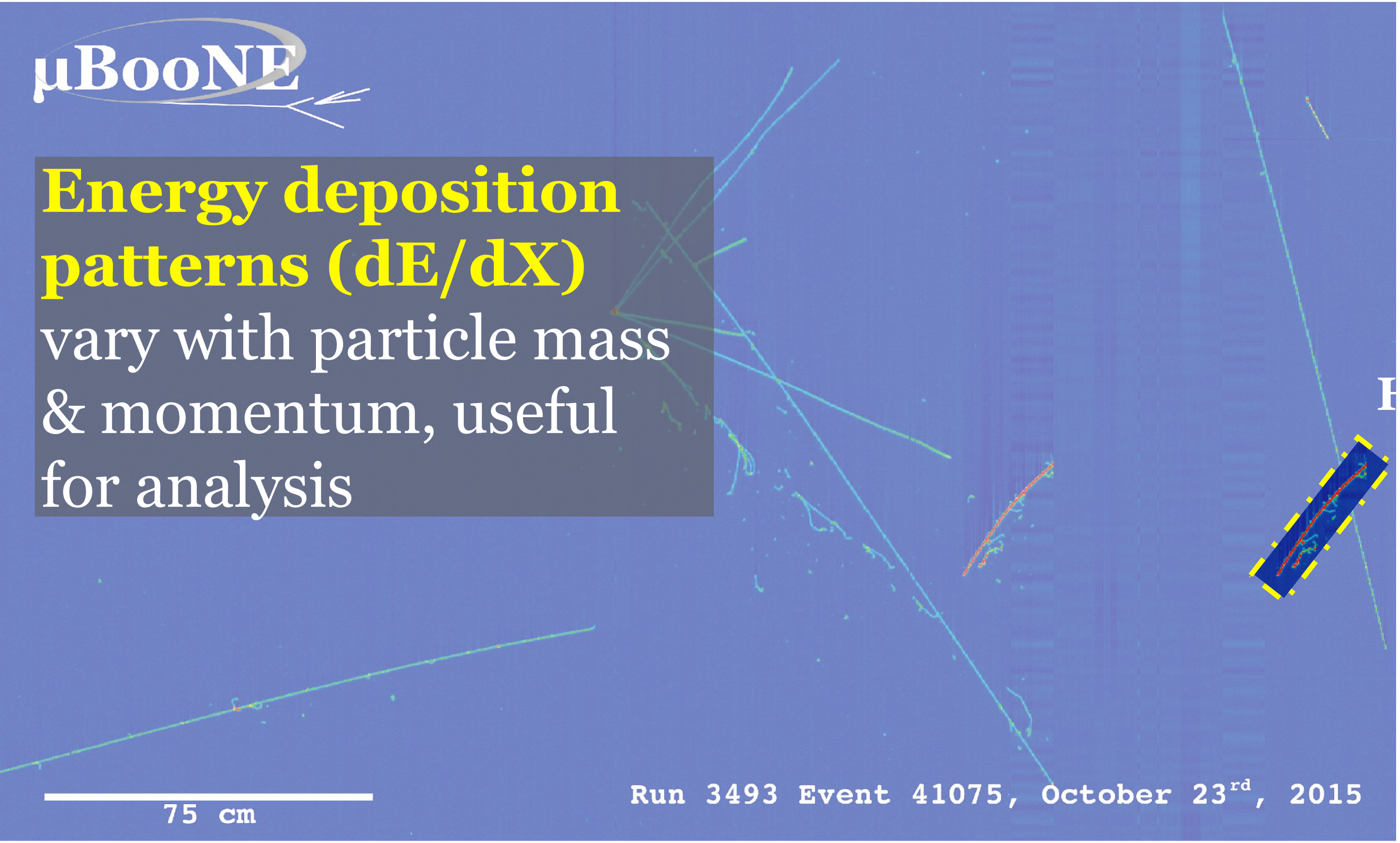


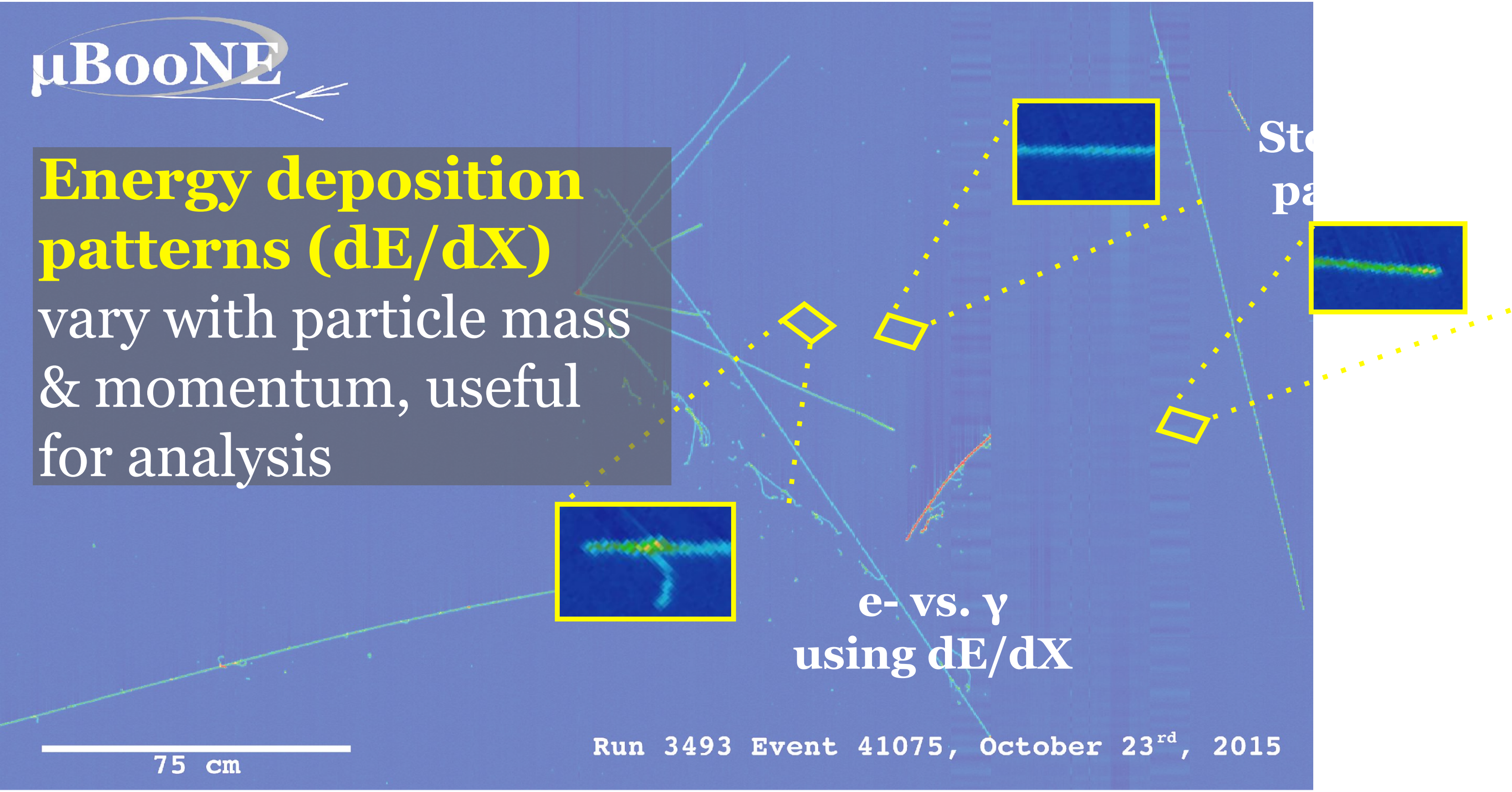
Many, local kinks caused by Multiple Coulomb Scattering process can be used for momentum estimation



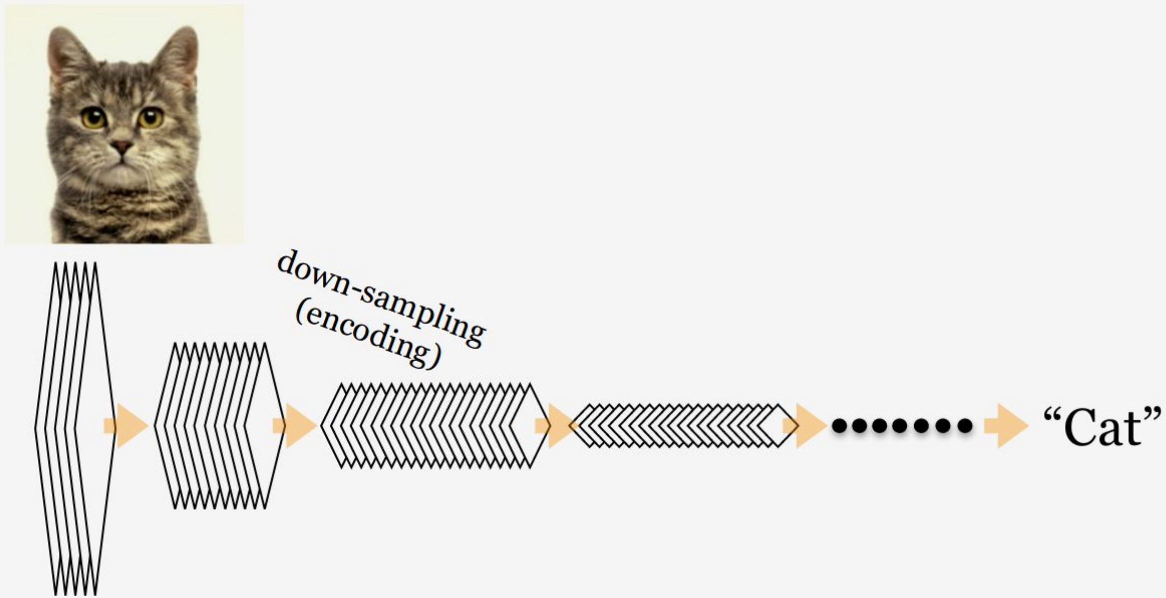
Small branches on muon-like trajectories are knocked-off electrons, useful key for the direction

Energy deposition patterns (dE/dX) vary with particle mass & momentum, useful for analysis

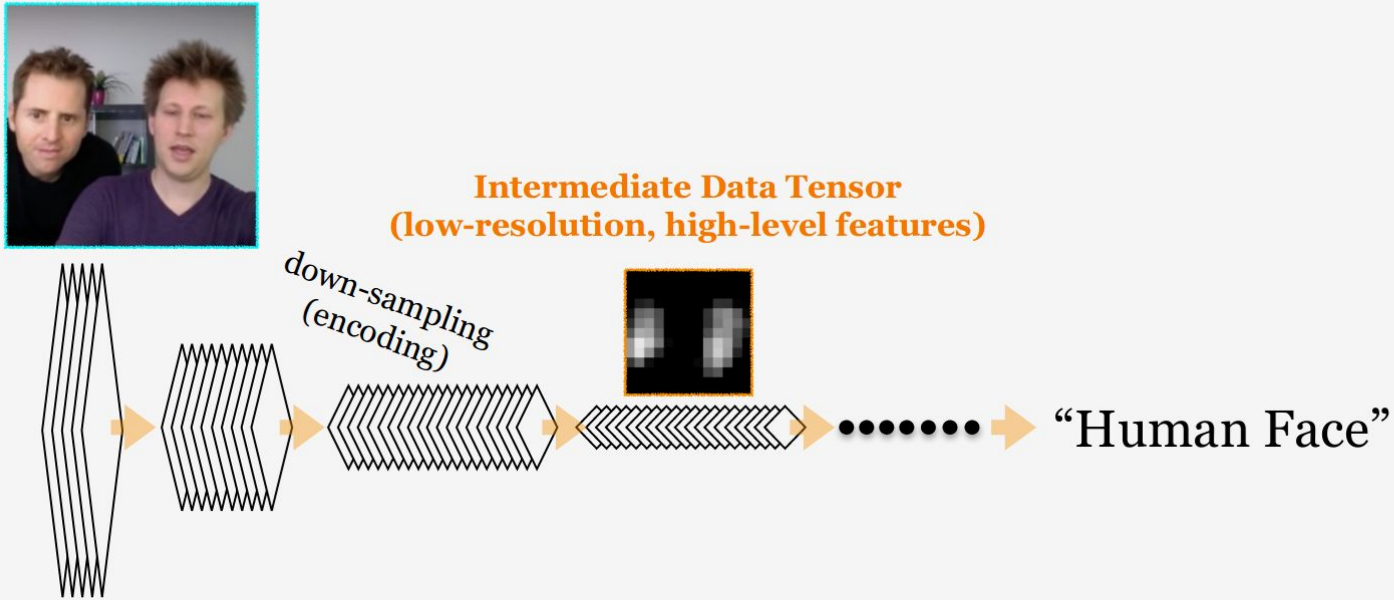



μBooNE
Energy deposition patterns (dE/dX)
vary with particle mass & momentum, useful for analysis
e- vs. γ
using dE/dX
75 cm
Run 3493 Event 41075, October 23rd, 2015

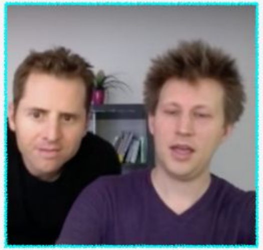
How image classification works



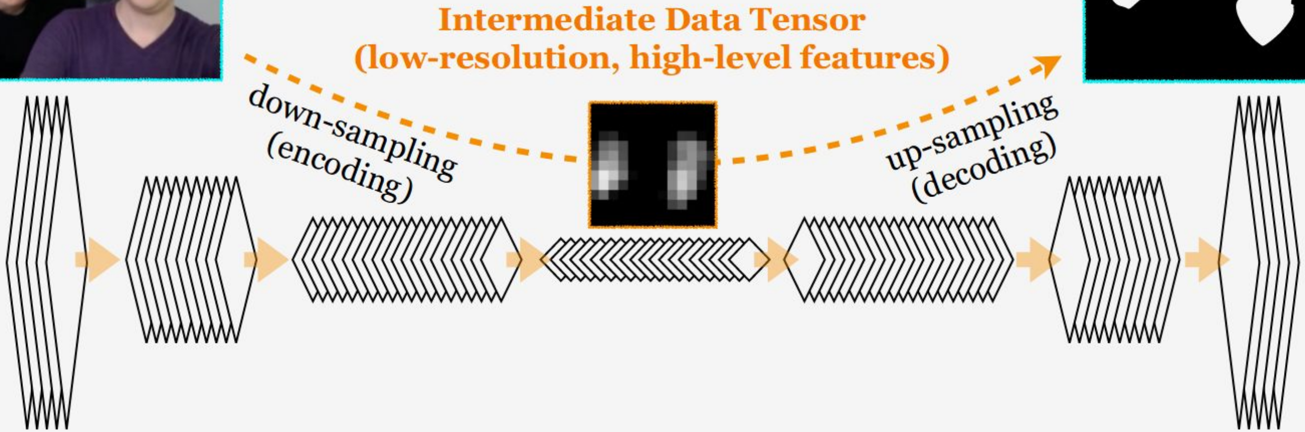
How image classification works



How pixel segmentation works

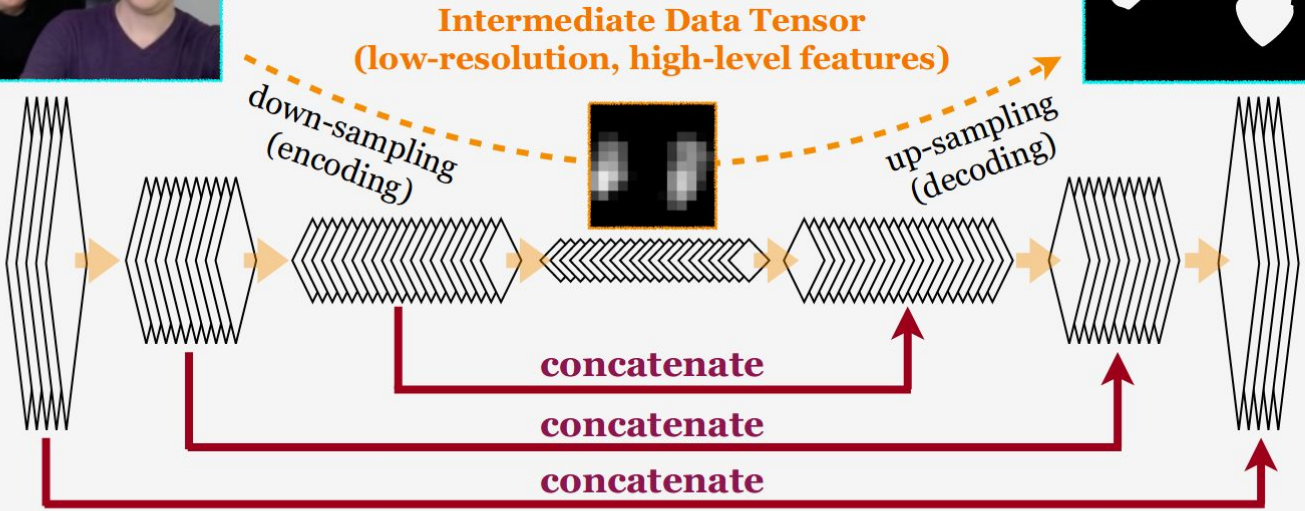


- Combine “up-sampling” + convolutions
- Output: “learnable” interpolation filters



How pixel segmentation works

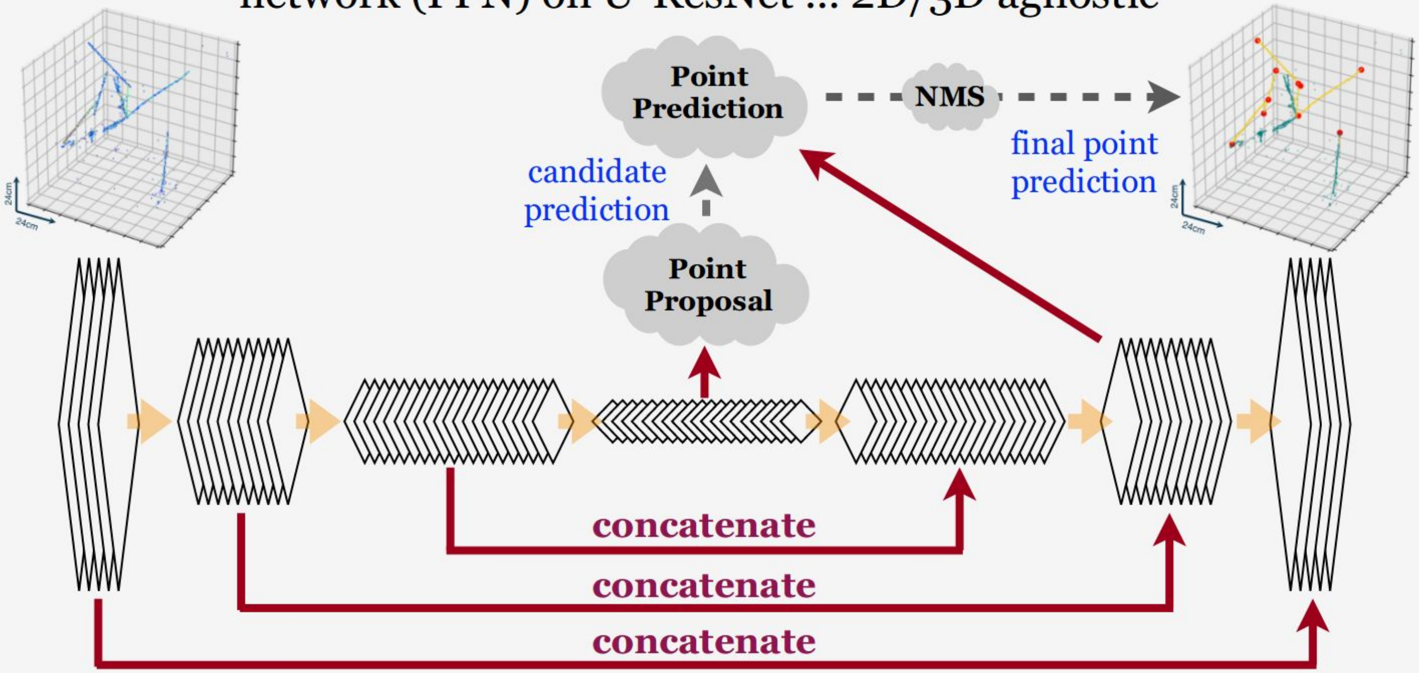
- Combine “up-sampling” + convolutions
- Output: “learnable” interpolation filters



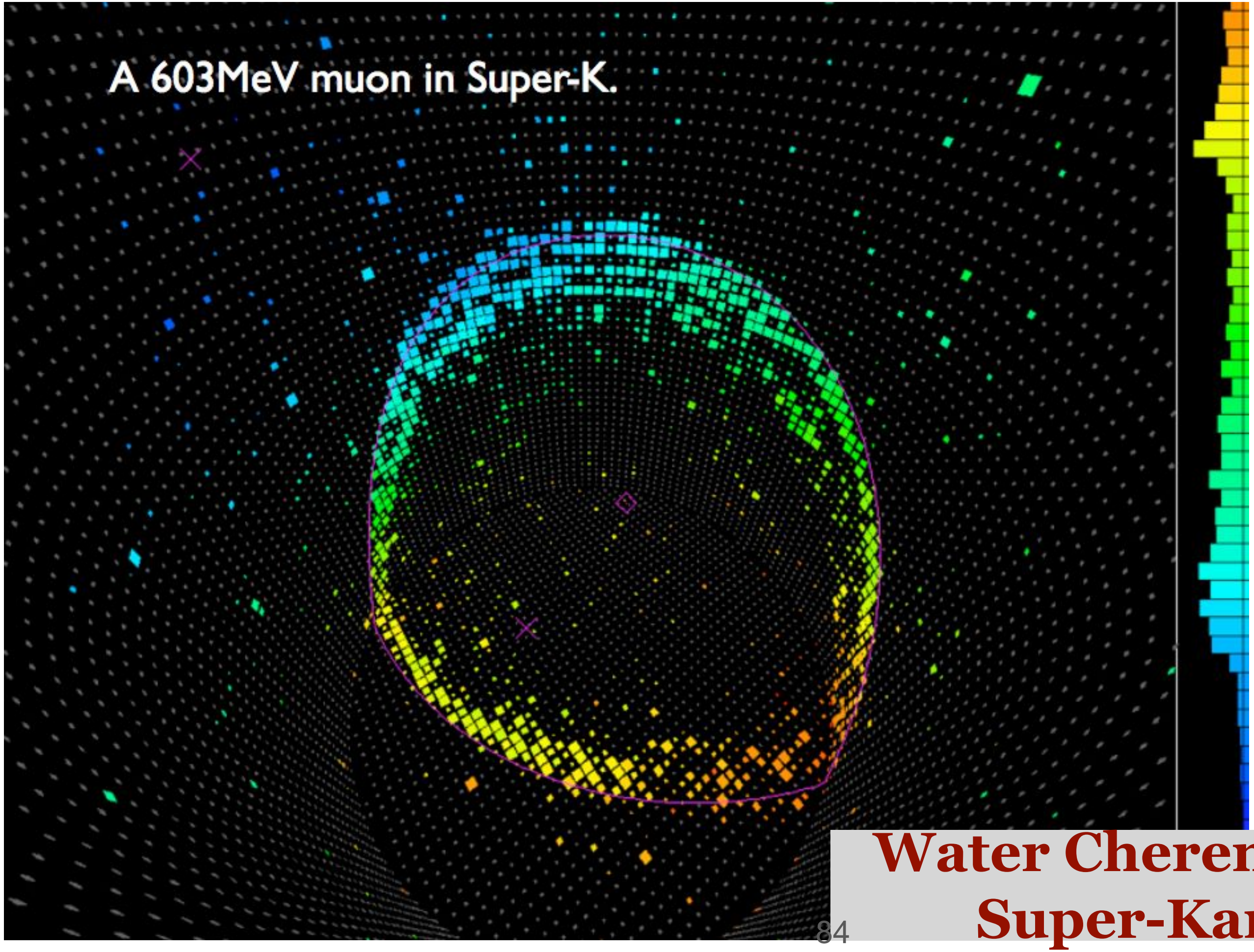
Concatenation recovers spatial resolution information

HO HO HO
Machine Learning for Particle Image Analysis

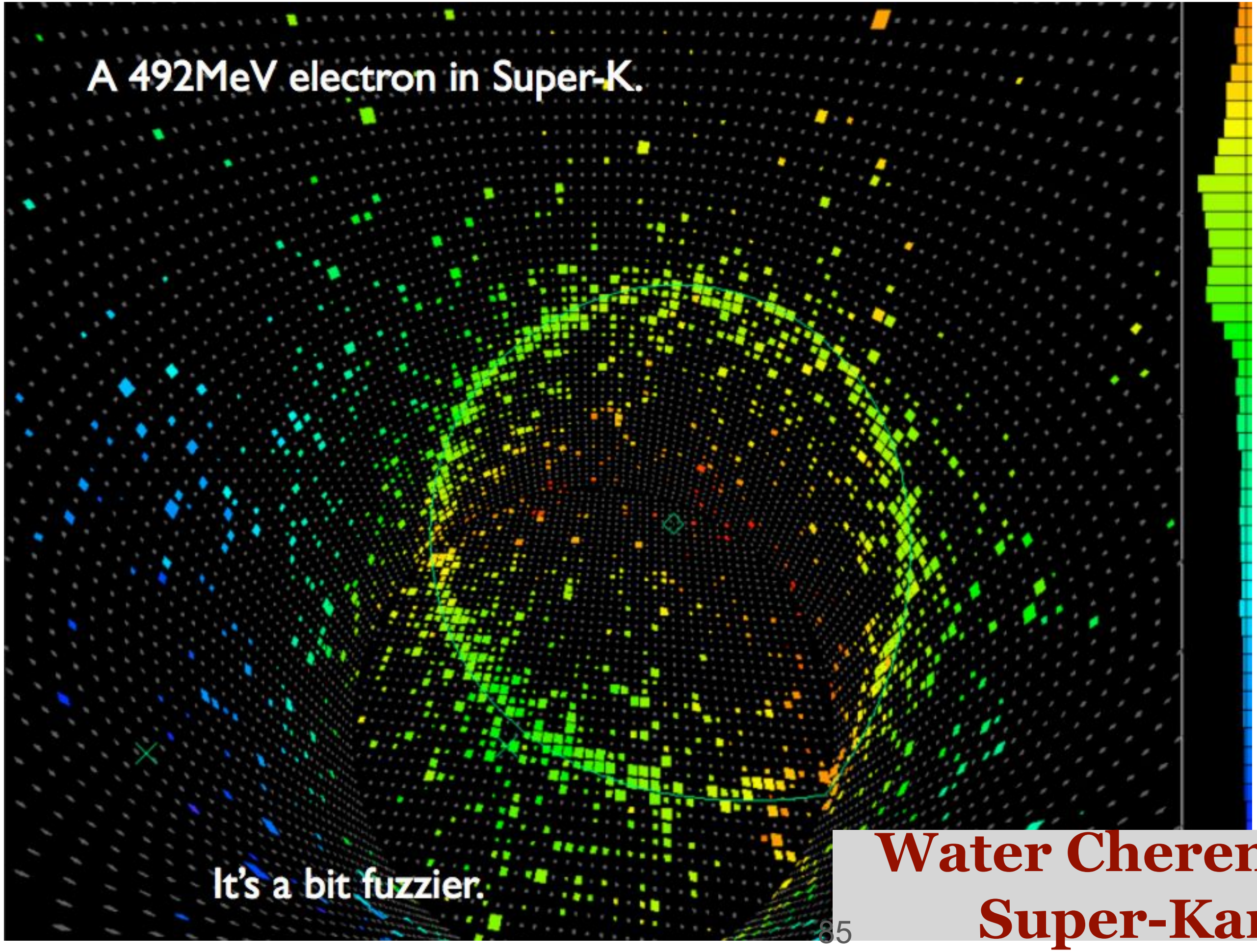
Parasitic multi-task scheme for point prediction network (PPN) on U-ResNet … 2D/3D agnostic



Concatenation recovers spatial resolution information

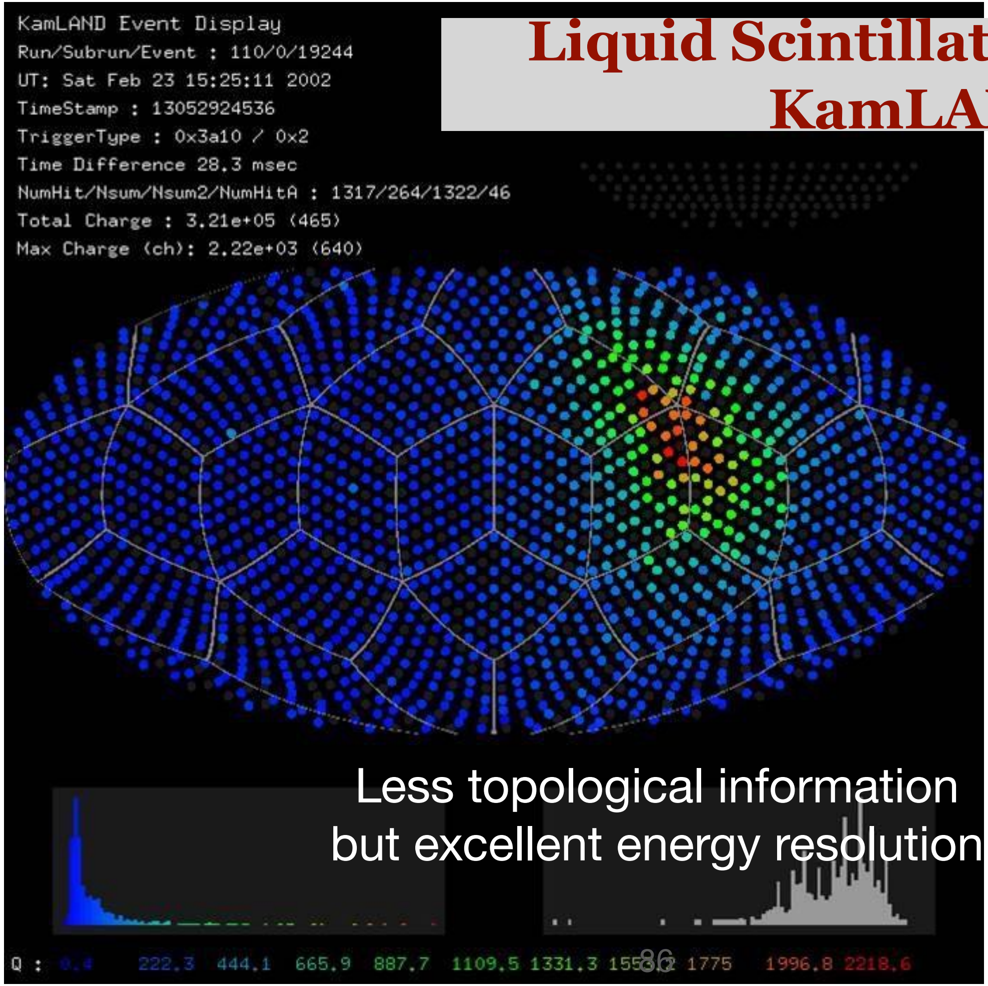


Water Cherenkov Detector
Super-Kamiokande



Water Cherenkov Detector
Super-Kamiokande

Liquid Scintillator Detector
KamLAND



Less topological information
but excellent energy resolution