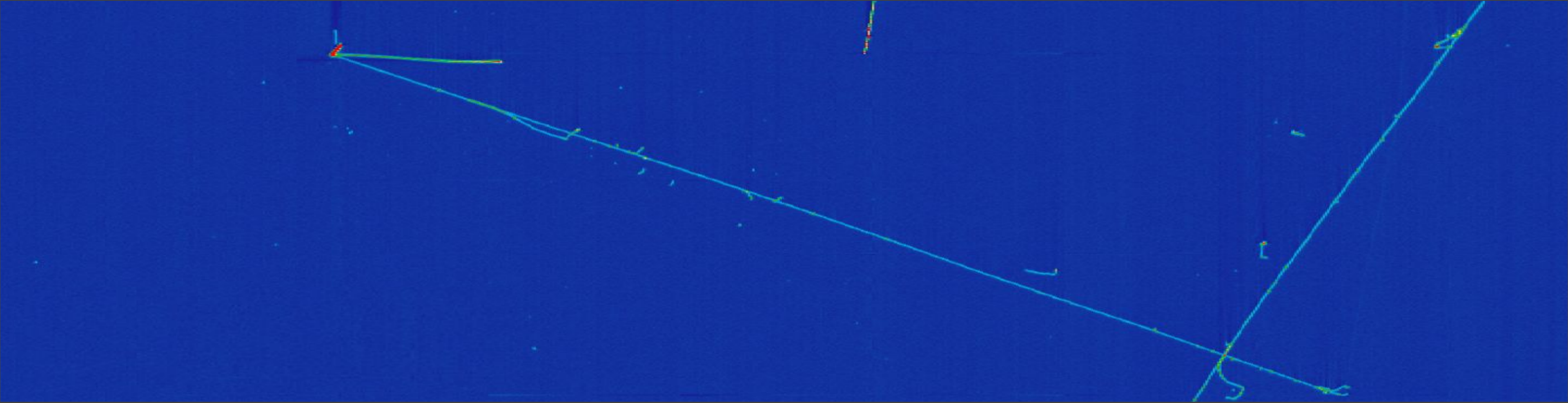


Machine Learning for Particle Imaging Detectors in Experimental Neutrino Physics

Kazuhiro Terao

SLAC National Accelerator Lab.

Machine Learning and Reconstruction for Neutrino Experiments (Sept. DESY)

U.S. DEPARTMENT OF ENERGY Office of Science

SLAC NATIONAL ACCELERATOR LABORATORY

About me

At SLAC, $$ supported by HEP + ML initiatives

- R&D of ML applications for physics experiments
  - "DL-based data reco for LArTPCs", "Hierarchical probabilistic and generative model to measure and constrain the impact of uncertainties", "Extreme scale particle tracking using distributed ML and graph neural network"
  - ATLAS/CMS/LSST/$0\nu\beta\beta$/Accelerator/Biology/Photon Sci. ...

Machine Learning & Computer Vision in Neutrino Physics

About me

At SLAC, $$ supported by HEP + ML initiatives

- R&D of ML applications for physics experiments
  - "DL-based data reco for LArTPCs", "Hierarchical probabilistic and generative model to measure and constrain the impact of uncertainties", "Extreme scale particle tracking using distributed ML and graph neural network"
  - ATLAS/CMS/LSST/$0\nu\beta\beta$/Accelerator/Biology/Photon Sci. ...

- Organize/Contribute ML workshops/summer schools
  - e.g): Hammer-and-Nails, ML-HEP summer school, Hyper-K Canada ML workshop, SLAC ML Crash Course, KMI Software School, CoDaS-HEP
  - **Interested in code-sprint to get something working?** Let me know! (ML, open data R&D + management, software container standardization and distribution, etc.)

About me

At SLAC, $$ supported by HEP + ML initiatives

Advertisement :)

Got cool results? Share with us at **SLAC AI seminar** (sure, we'll pay your travel!)

Got interest in ML-for-physics?
Apply for **a job at SLAC ML-initiatives**!

Got interest in ML workshop?
Let's organize together!

Neutrino + ML post-doc? Contact me :)

ments
ical probabilistic
in the impact of
g using

y/Photon Sci. ...

ner schools
chool, Hyper-K
e, KMI Software

ing working?
ment, software
container standardization and distribution, etc.)

About this talk

SLAC

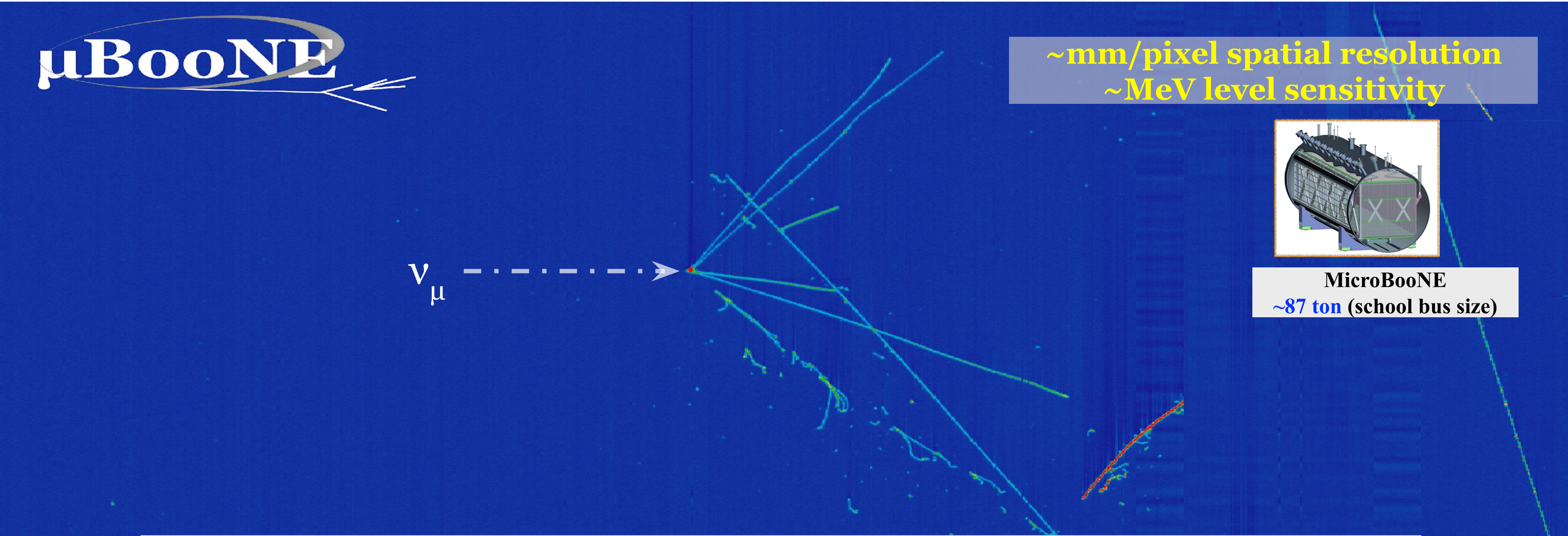
Today: ML-based data **reconstruction** for LArTPCs

Outline:

1. Why machine learning? Why reconstruction?
2. Reconstruction technique “zoo”
   a. Object detection & semantic segmentation
   b. Sparse convolutions toward 3D data
   c. Clustering using CNN and GNN
3. Softwares: container, open data, and distributed ML

Time Projection Chambers

µBooNE



~mm/pixel spatial resolution
~MeV level sensitivity

MicroBooNE
~87 ton (school bus size)

Liquid Argon Time Projection Chamber

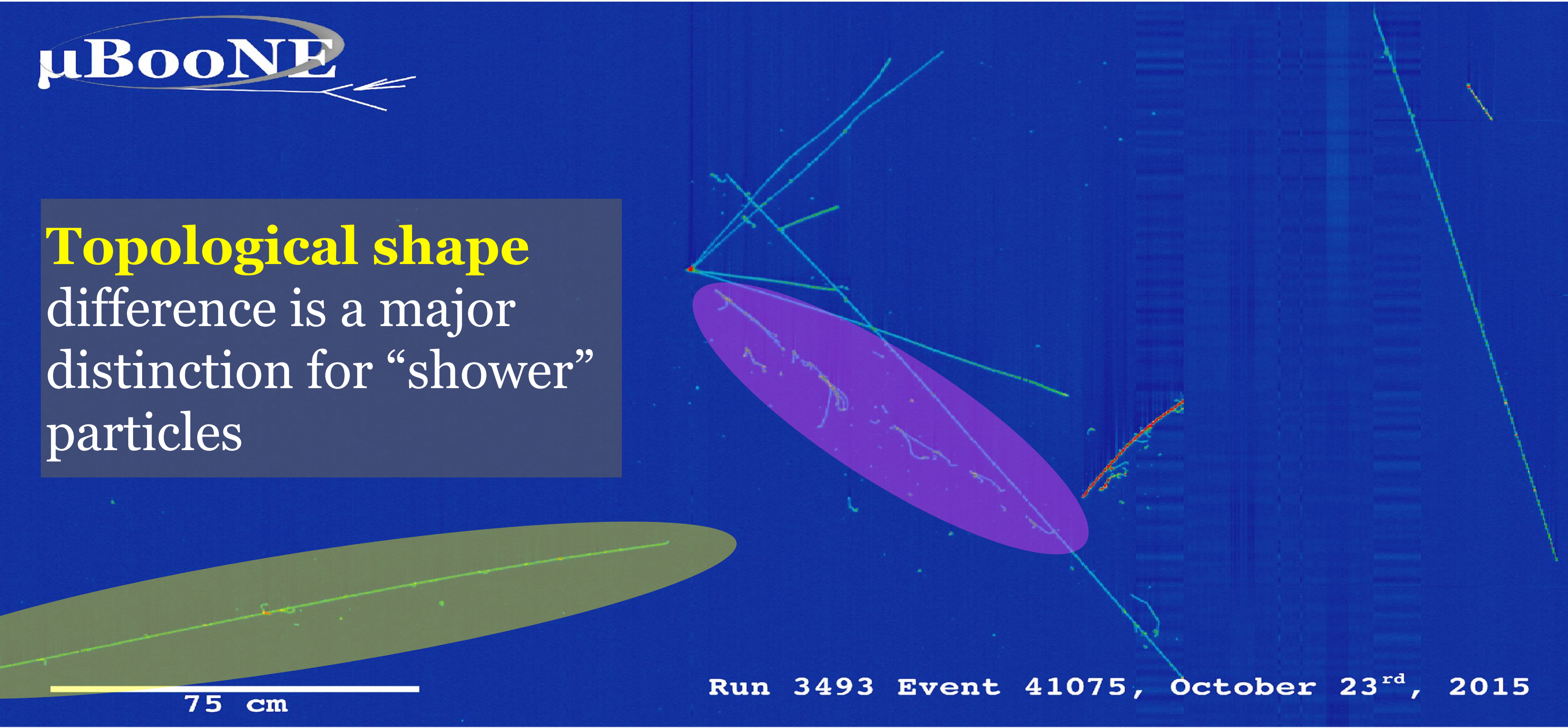
- High resolution photograph of charged particle trajectories
- Calorimetric measurement + scalability to a large mass

Machine Learning & Computer Vision in Neutrino Physics

Time Projection Chambers

SLAC

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



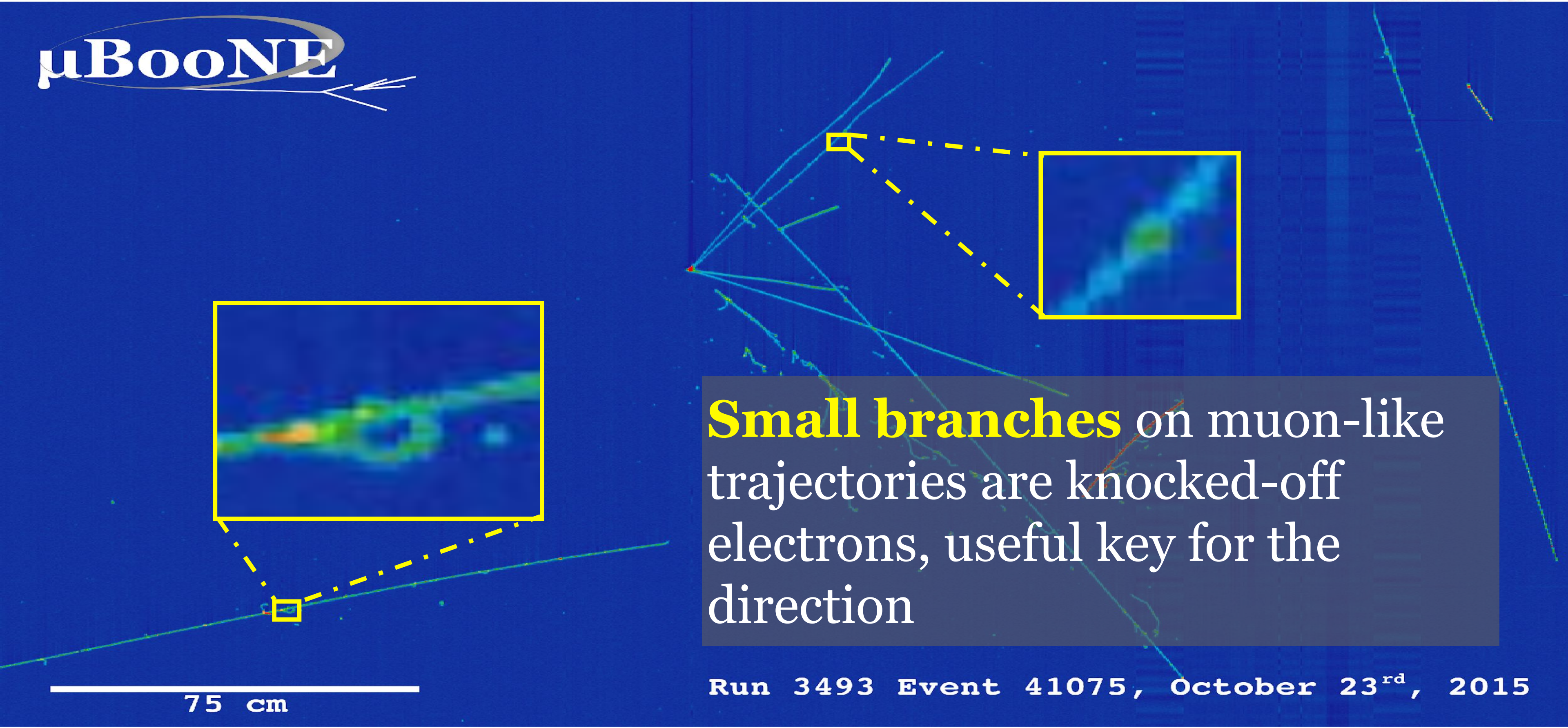
Machine Learning & Computer Vision in Neutrino Physics
Time Projection Chambers

SLAC

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

Machine Learning & Computer Vision in Neutrino Physics

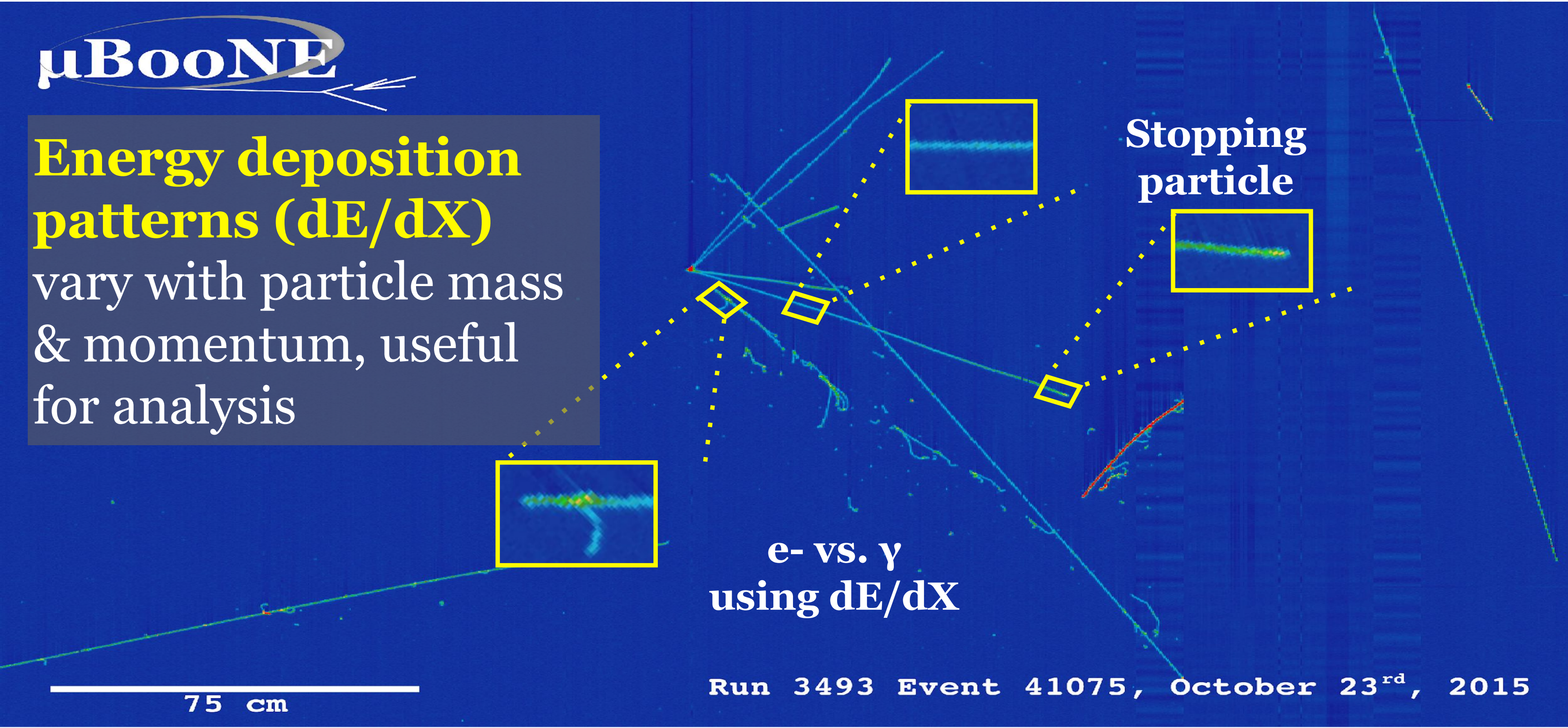
Time Projection Chambers



Machine Learning & Computer Vision in Neutrino Physics

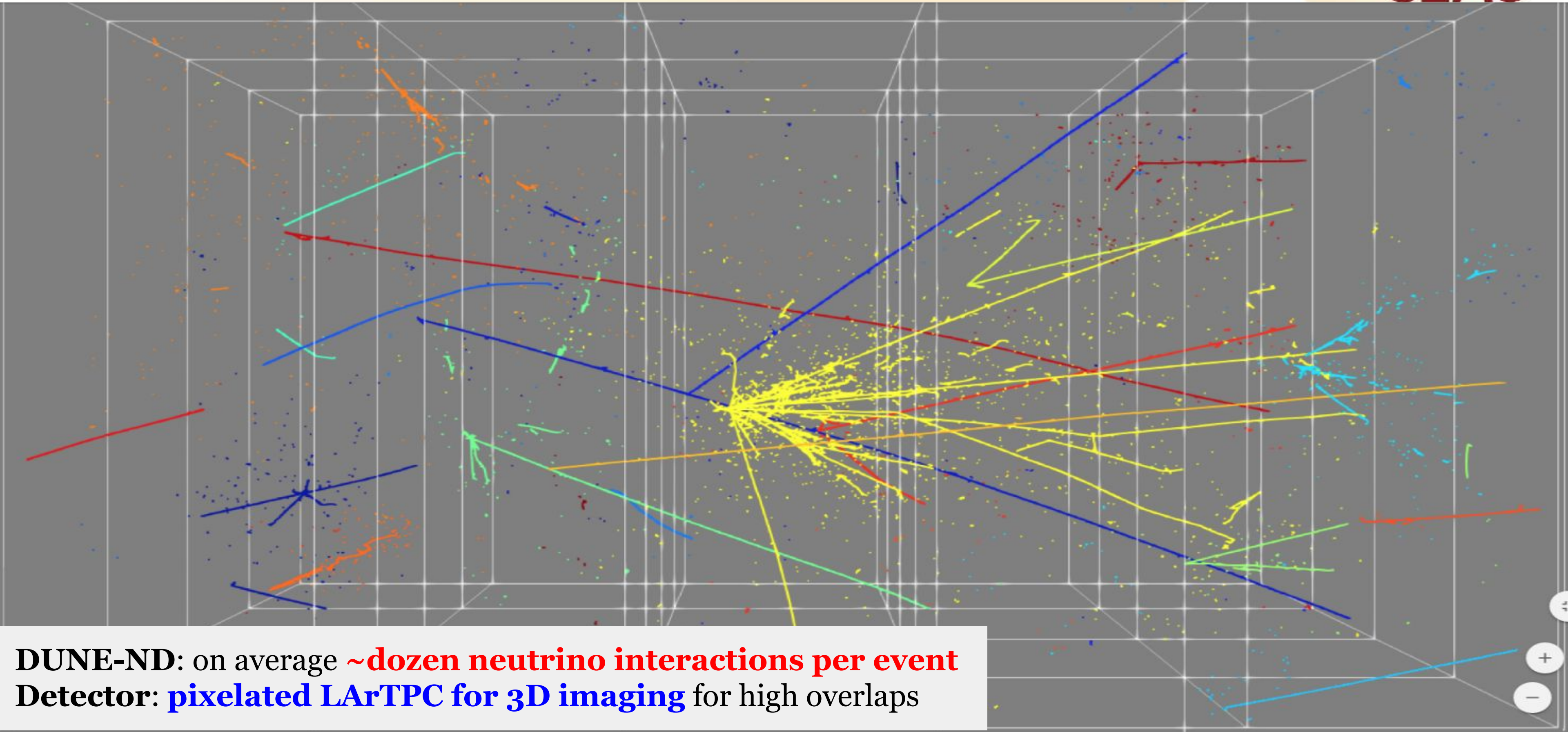
Time Projection Chambers

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



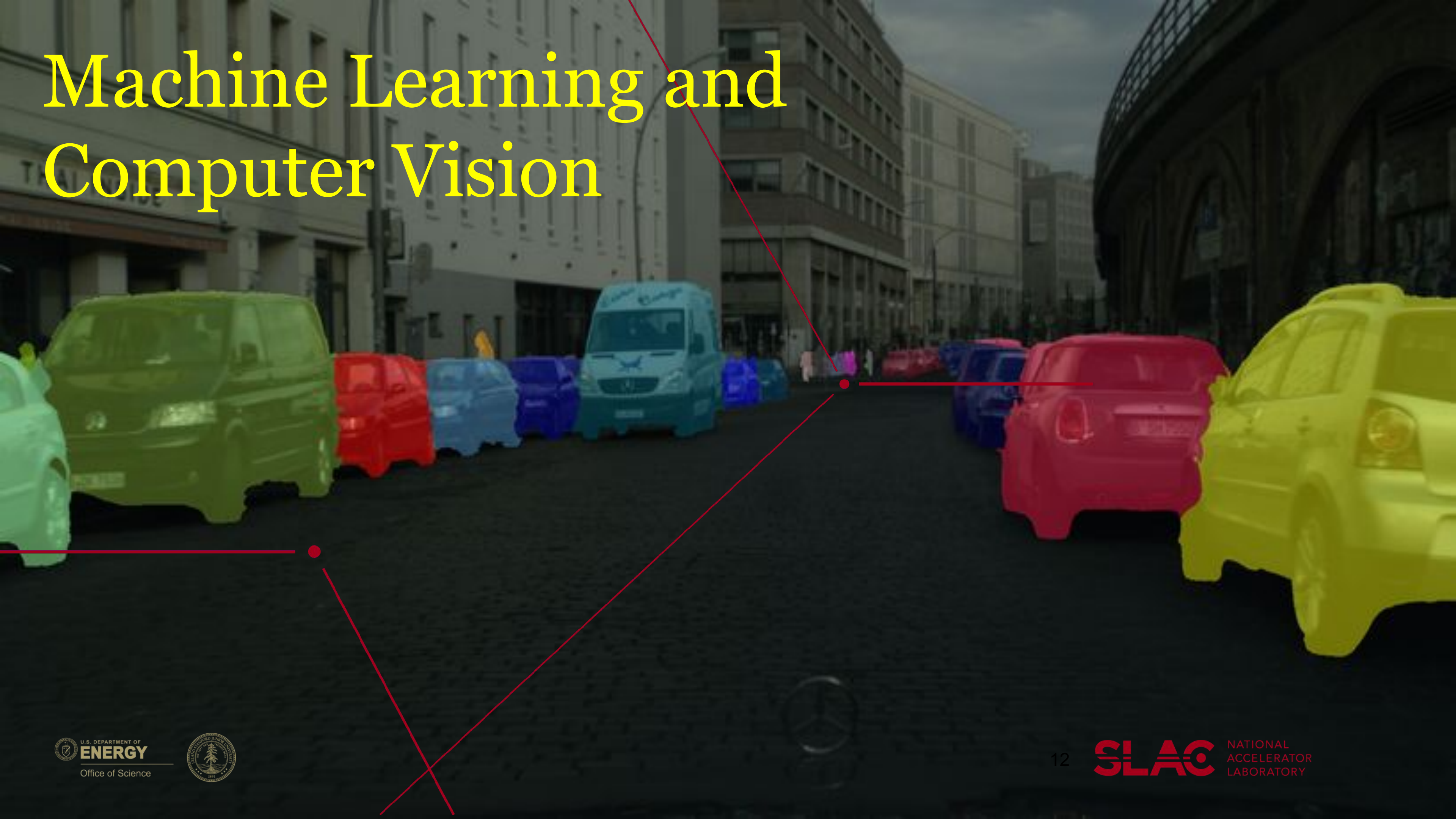
Machine Learning & Computer Vision in Neutrino Physics

Time Projection Chambers (3D ones)



DUNE-ND: on average **~dozen neutrino interactions per event**
Detector: **pixelated LArTPC for 3D imaging** for high overlaps

Machine Learning and Computer Vision



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

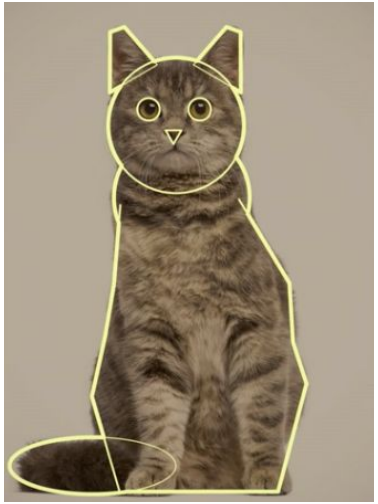
How to write an algorithm to identify a cat?

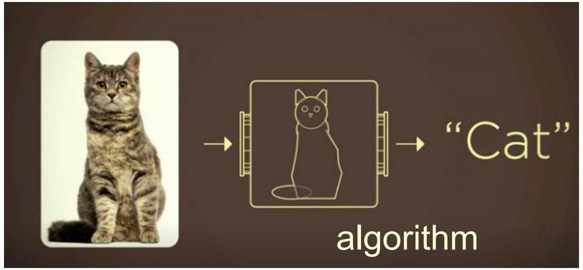
… very hard task …

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





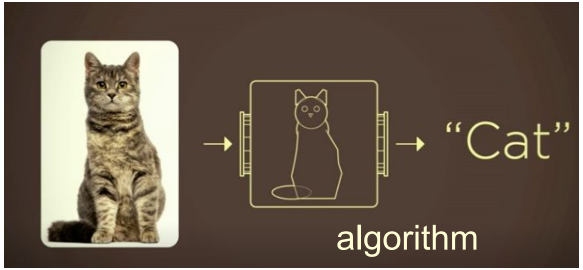
A cat = collection of certain shapes
(or, a neutrino)

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)

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.

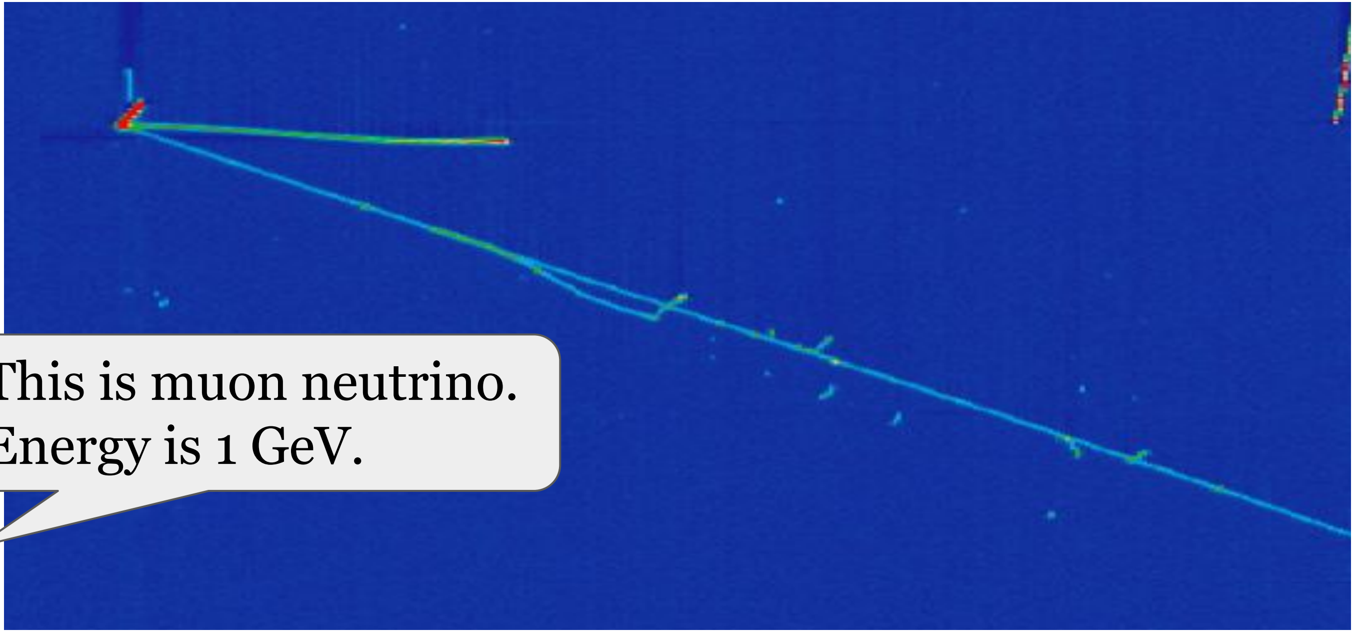
"Machine learning"

- Automatization of step 2, 3, and 4.
- Well-defined error propagation (step 5).
- Can optimize the whole chain for physics.

Next: what kind of ML algorithms?

Why Data Reconstruction

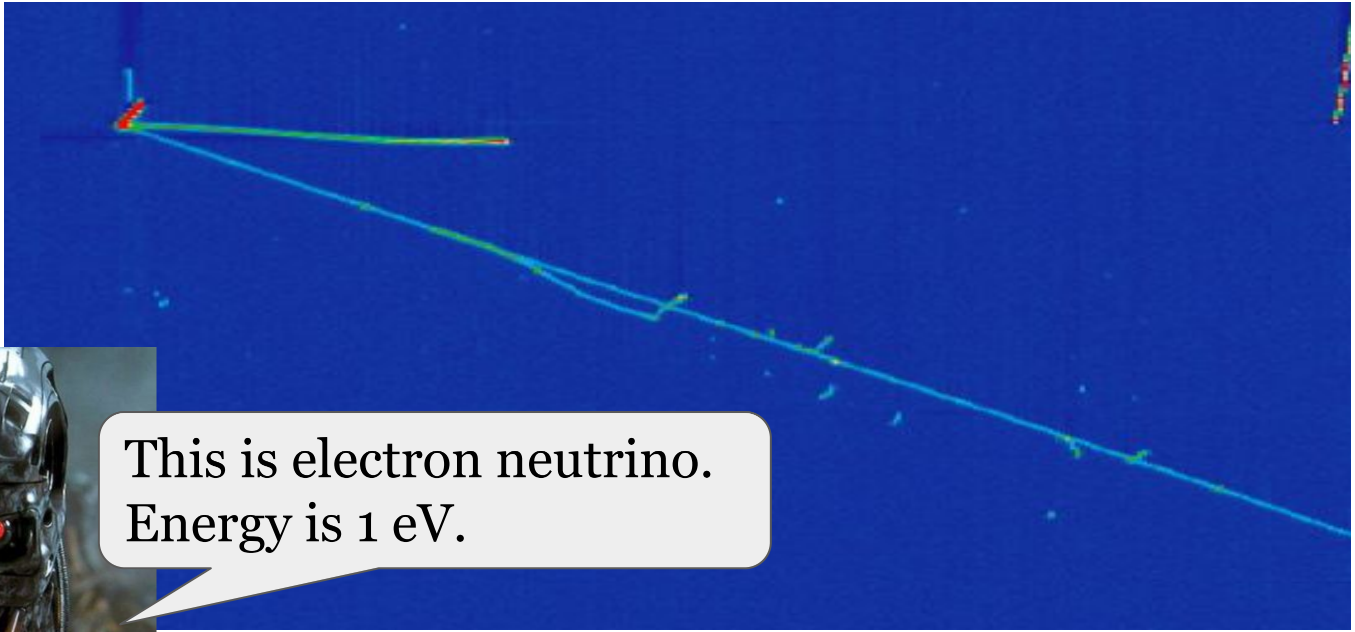
Image classification/regression: straight to “**flavour & energy**”



This is muon neutrino.
Energy is 1 GeV.

Machine Learning & Computer Vision in Neutrino Physics
Why Data Reconstruction

… **but also challenging:** a huge single-step of information reduction

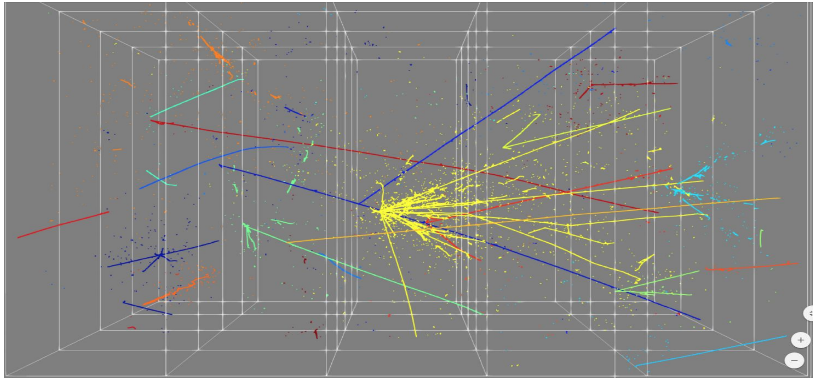


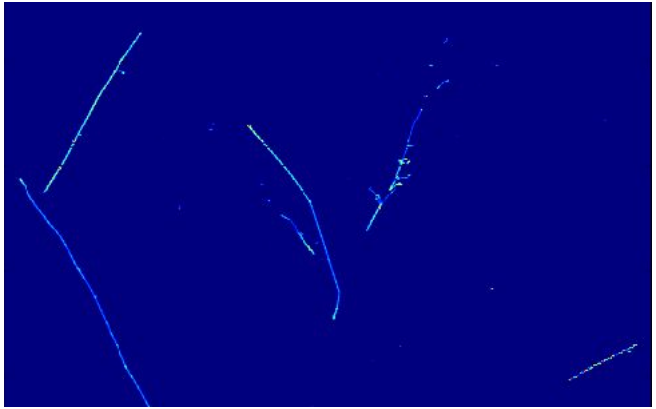
… **would be nice to know why you thought so …**

Why Data Reconstruction

SLAC

Also … **how do we get a "neutrino image"?**





… most of LArTPC detectors do not make "neutrino only" image …

- MicroBooNE, ICARUS, SBND, ProtoDUNE … physics in next 5 years
- DUNE-ND (a dozen neutrino interaction pile-up per event)

Machine Learning & Computer Vision in Neutrino Physics
Why Data Reconstruction

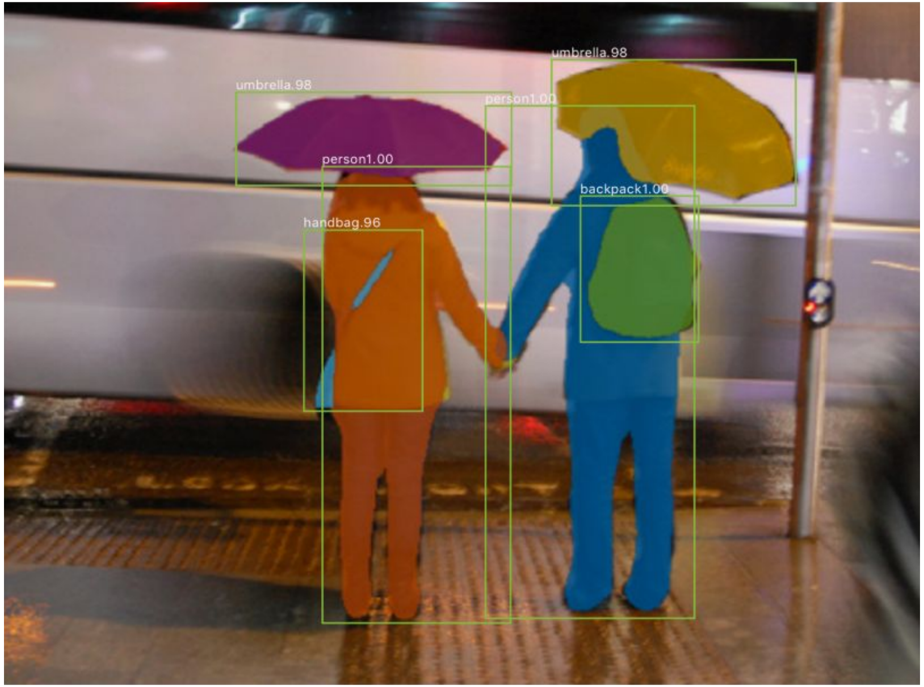


Image Context Identification

Why Data Reconstruction

"girl in pink dress is jumping in air."

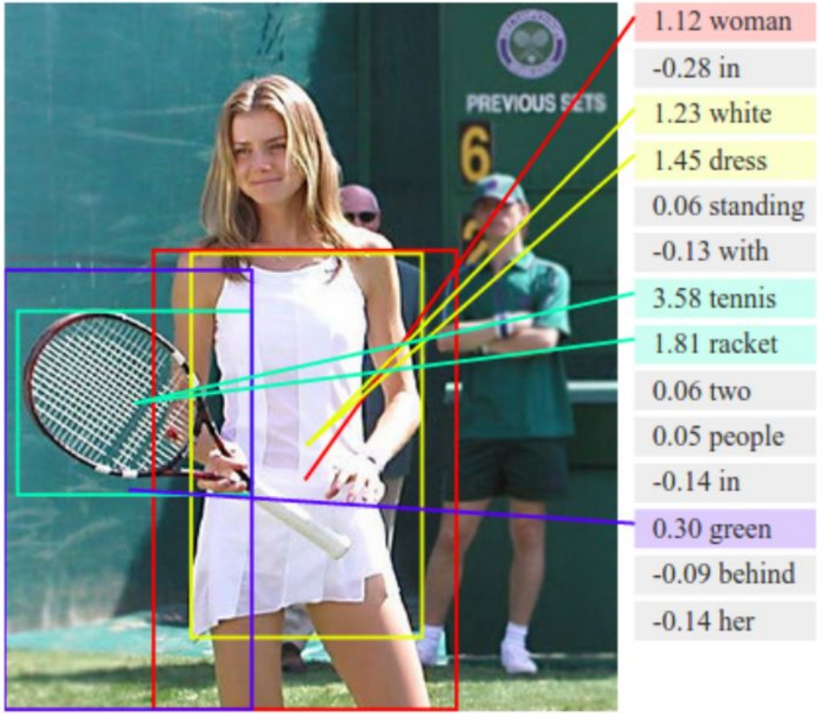


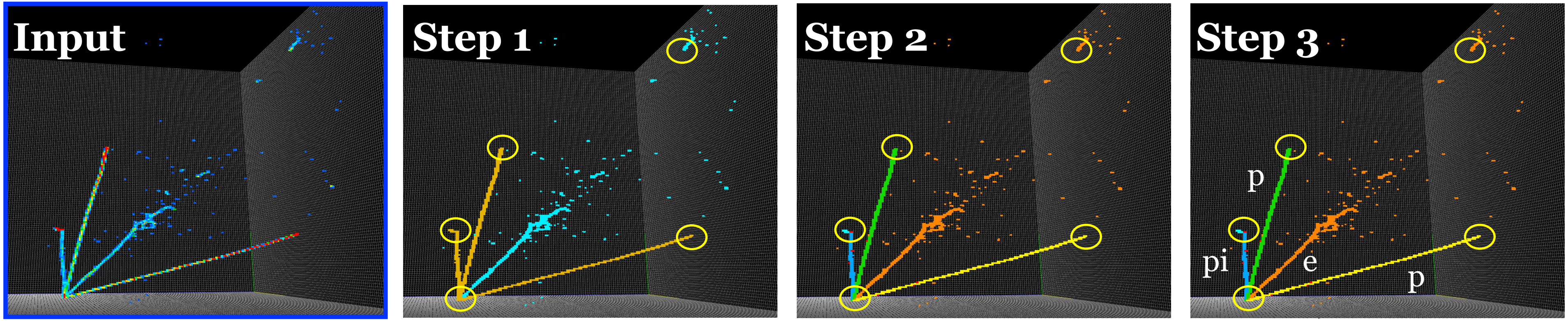
Image Context Correlation/Hierarchy Analysis

Data Reconstruction Chain

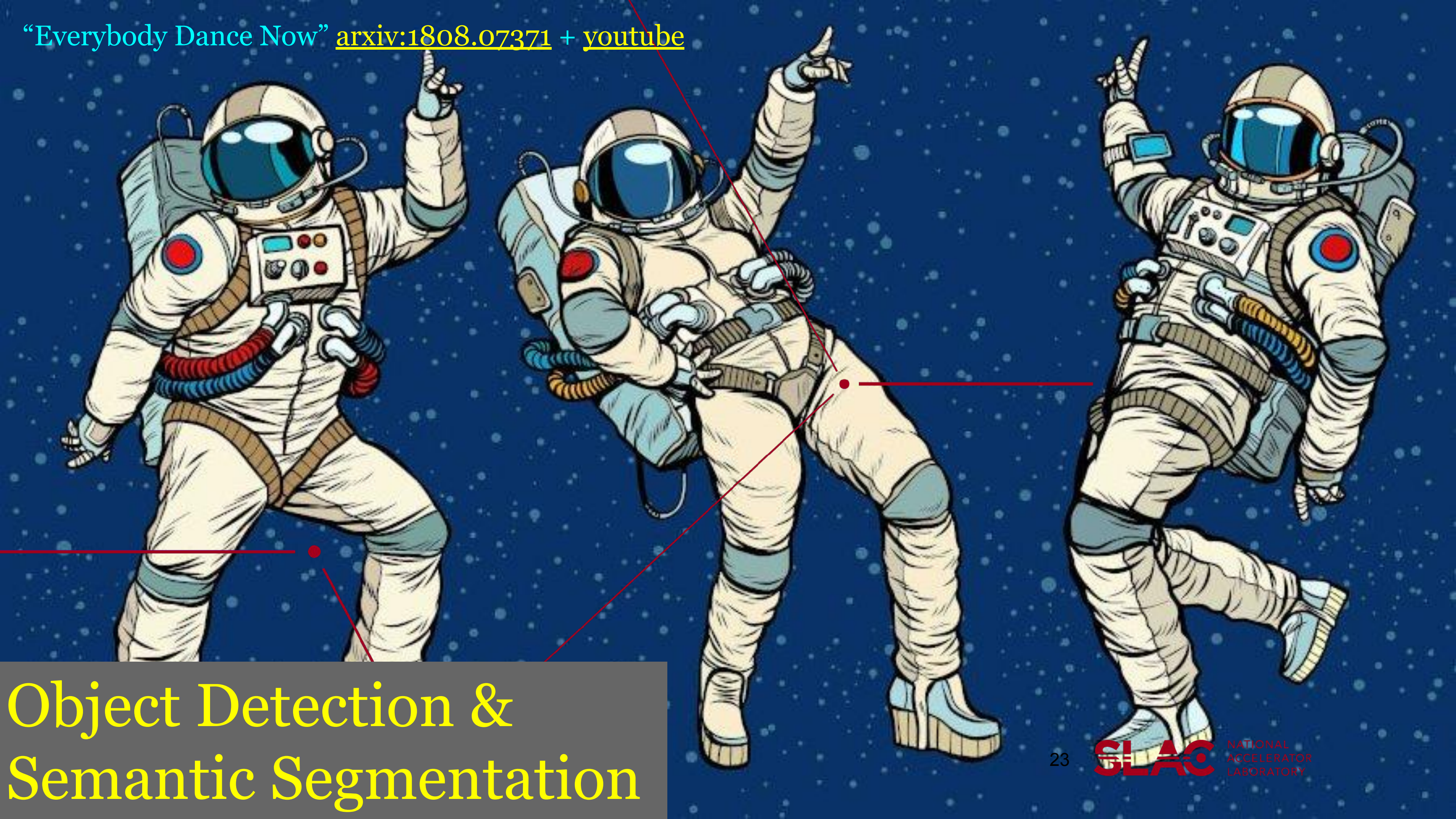
Extraction of hierarchical physics features…

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

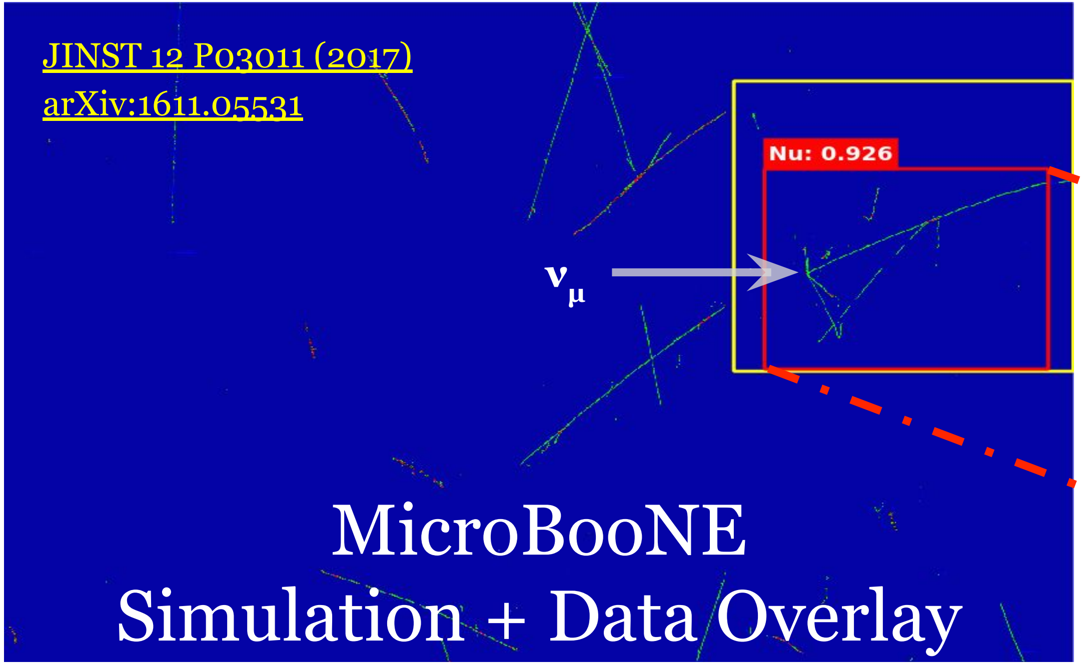


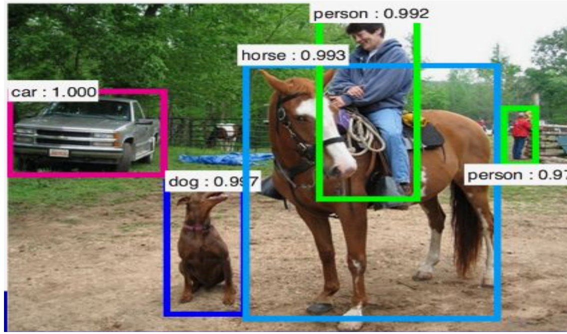
“Everybody Dance Now” arxiv:1808.07371 + youtube

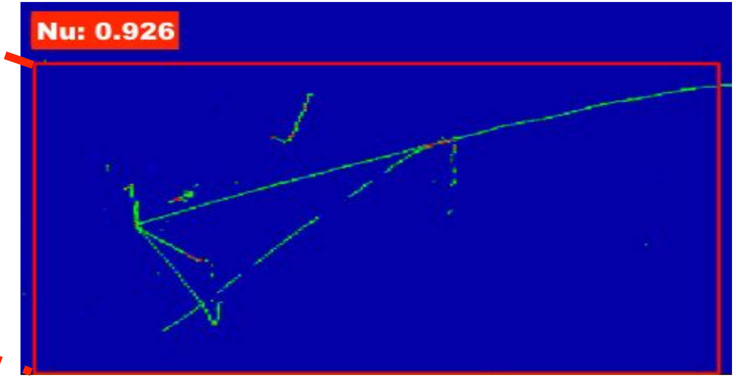


Object Detection & Semantic Segmentation

Neutrino Detection w/ R-CNN
(MicroBooNE LArTPC)





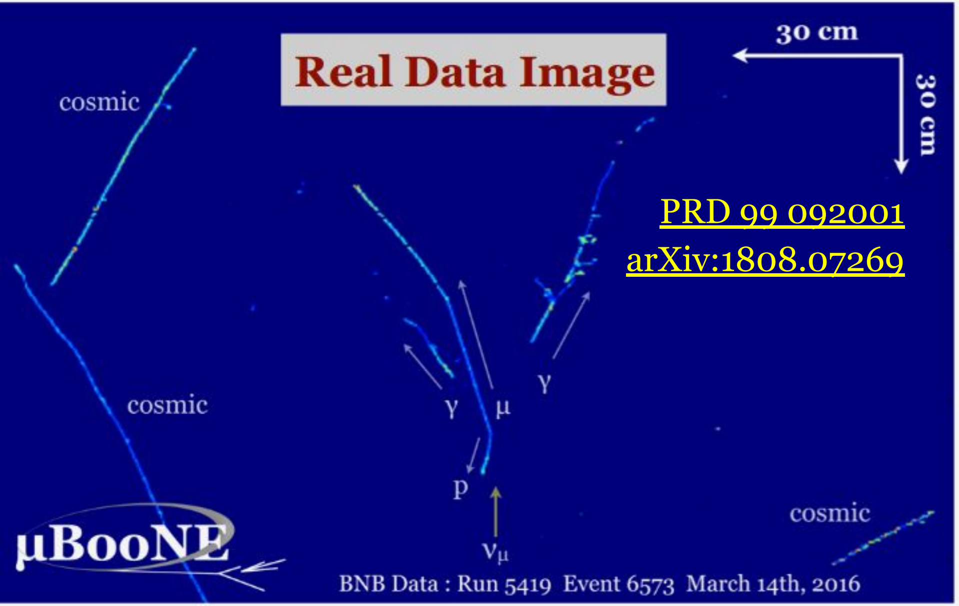


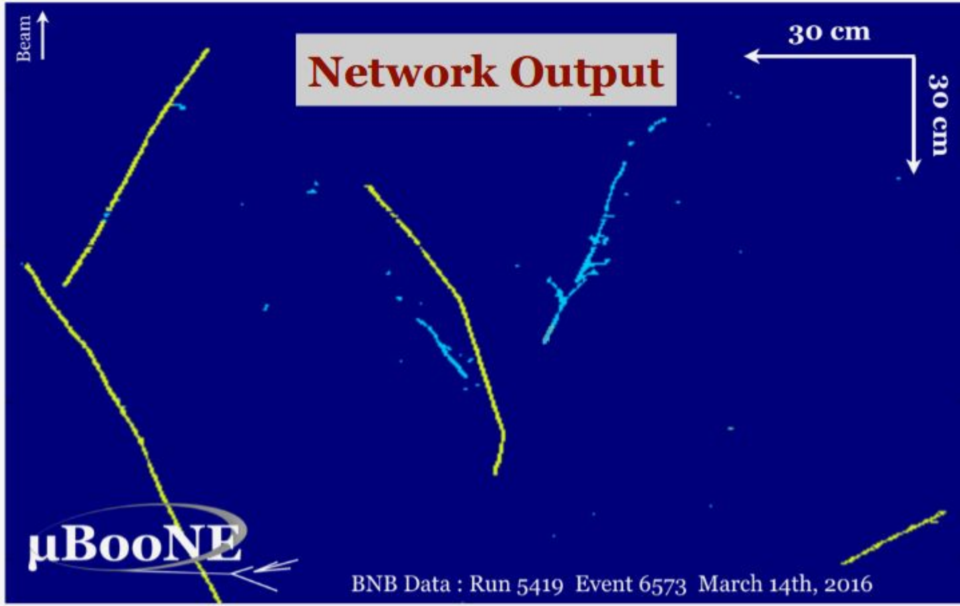
Task: propose a rectangular box that contains neutrino interaction (location & size)

Machine Learning & Computer Vision in Neutrino Physics

Semantic Segmentation for Pixel-level Particle ID

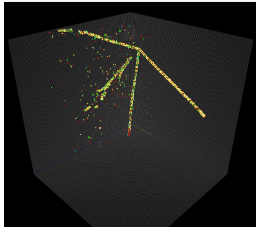
Separate electron/positron energy depositions from other types at raw waveform level.
Helps the downstream clustering algorithms (**data/sim comp. @ arxiv:1808.07269**)

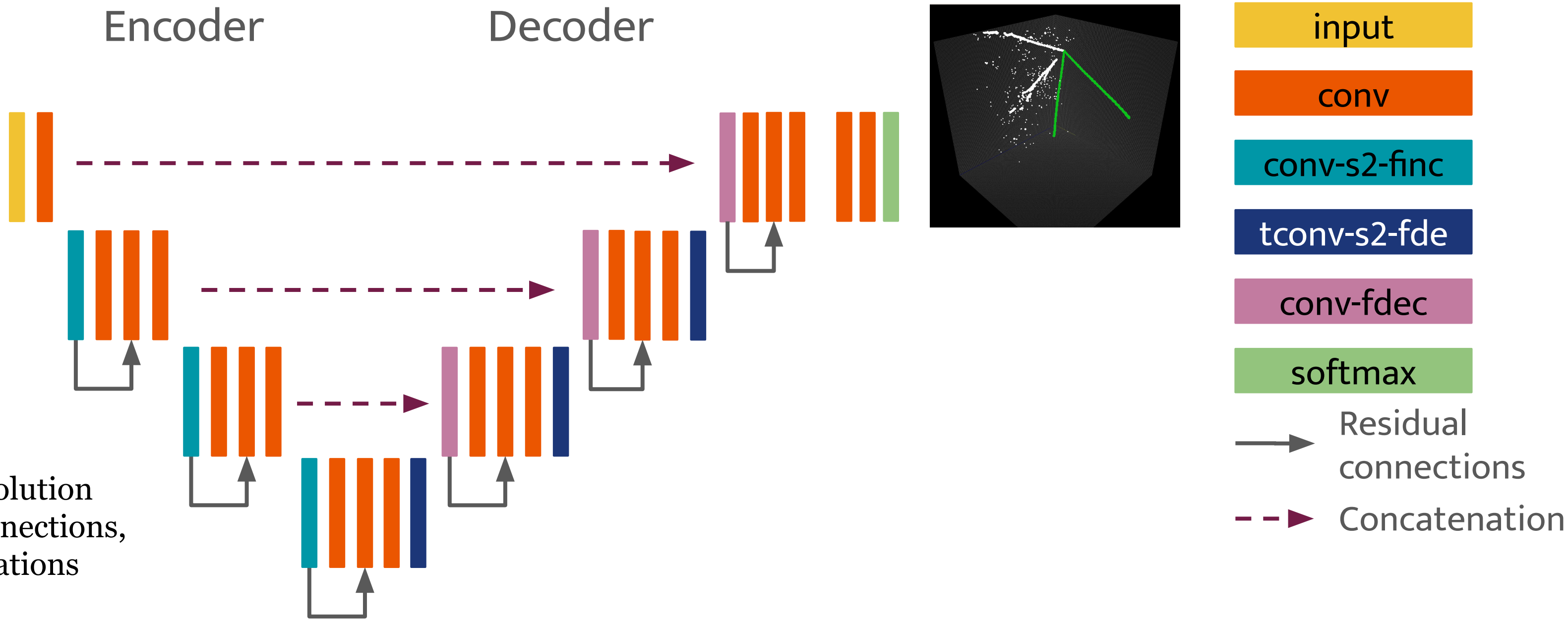




Network Input **Network Output**

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





Number of strided convolutions, convolution layers, residual connections, differ in impementations

Image credit: Laura Domine @ Stanford

Scalable CNN for Sparse Particle Imaging Data

"Applying CNN" is simple, but is it scalable for us?

LArTPC data is generally sparse, but locally dense

CNN applies **dense matrix operations**

In photographs, **all pixels are meaningful**

grey pixels = dolphines, blue pixels = water, etc…

Figures/Texts: courtesy of Laura Domine @ Stanford

Scalable CNN for Sparse Particle Imaging Data

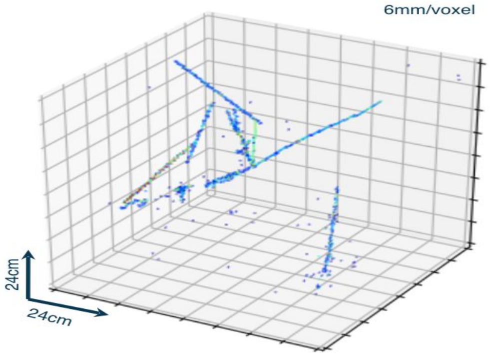
"Applying CNN" is simple, but **is it scalable for us?**

LArTPC data is generally sparse, but locally dense

CNN applies **dense matrix operations**

In photographs, **all pixels are meaningful**

grey pixels = dolphines, blue pixels = water, etc...



Empty pixels = no energy

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

Zero pixels are meaningless!

Figures/Texts: courtesy of Laura Domine @ Stanford

Figure credit: Laura Domine @ Stanford

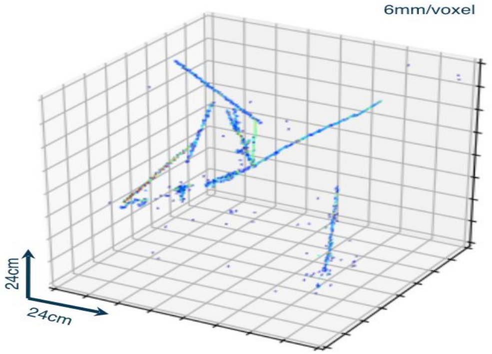
Scalable CNN for Sparse Particle Imaging Data

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

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 pixels does not

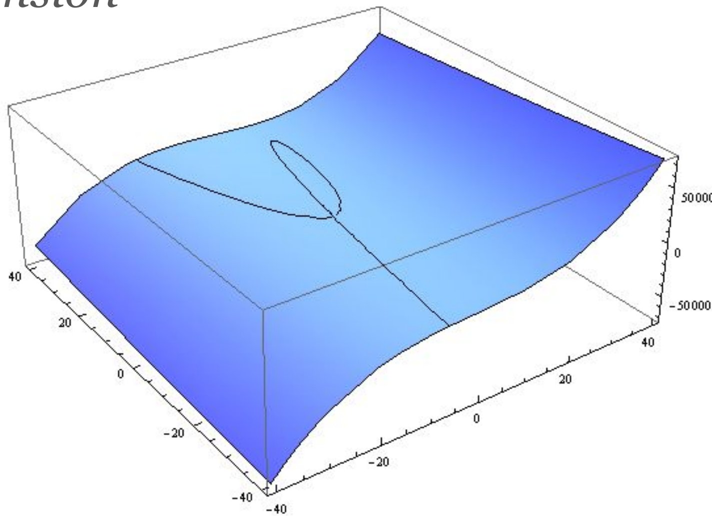
Figure credit: Laura Domine @ Stanford

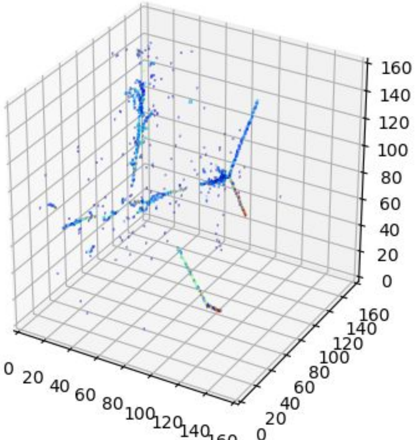
Submanifold Sparse Convolutions

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**...





Scalable CNN for Sparse Particle Imaging Data

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

... which makes convolution filter more effective on our data as long as the sparsity issue is handled

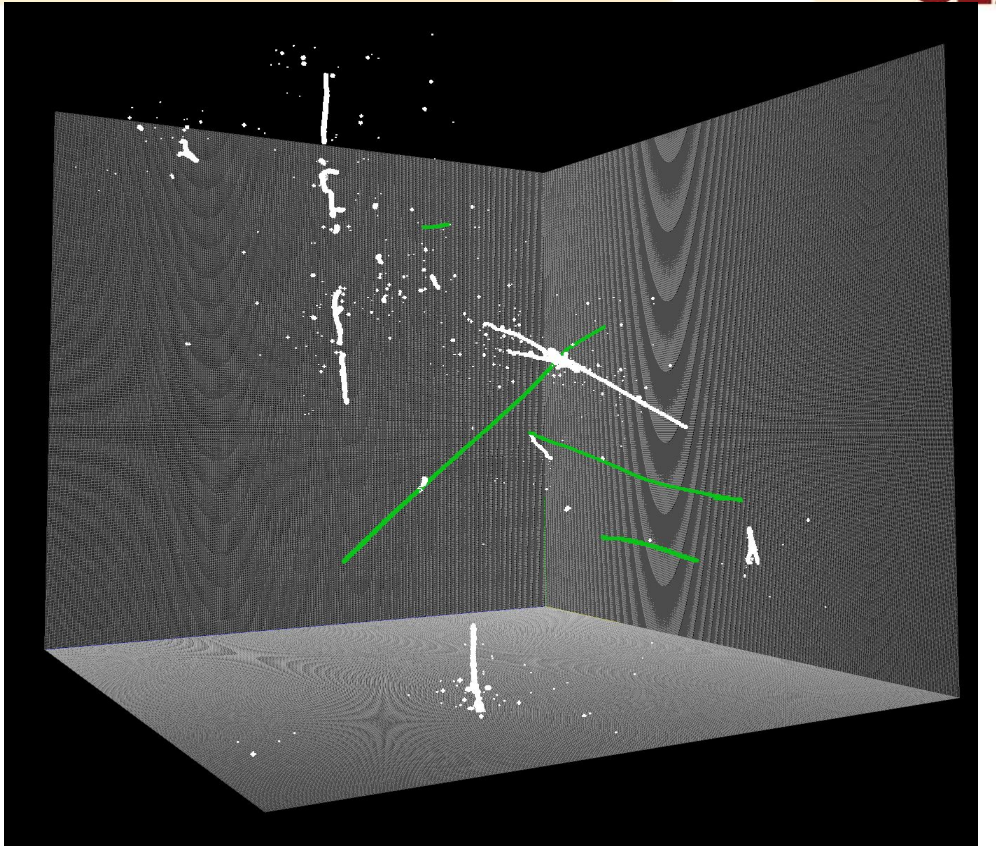


Image credit: Laura Domine @ Stanford

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

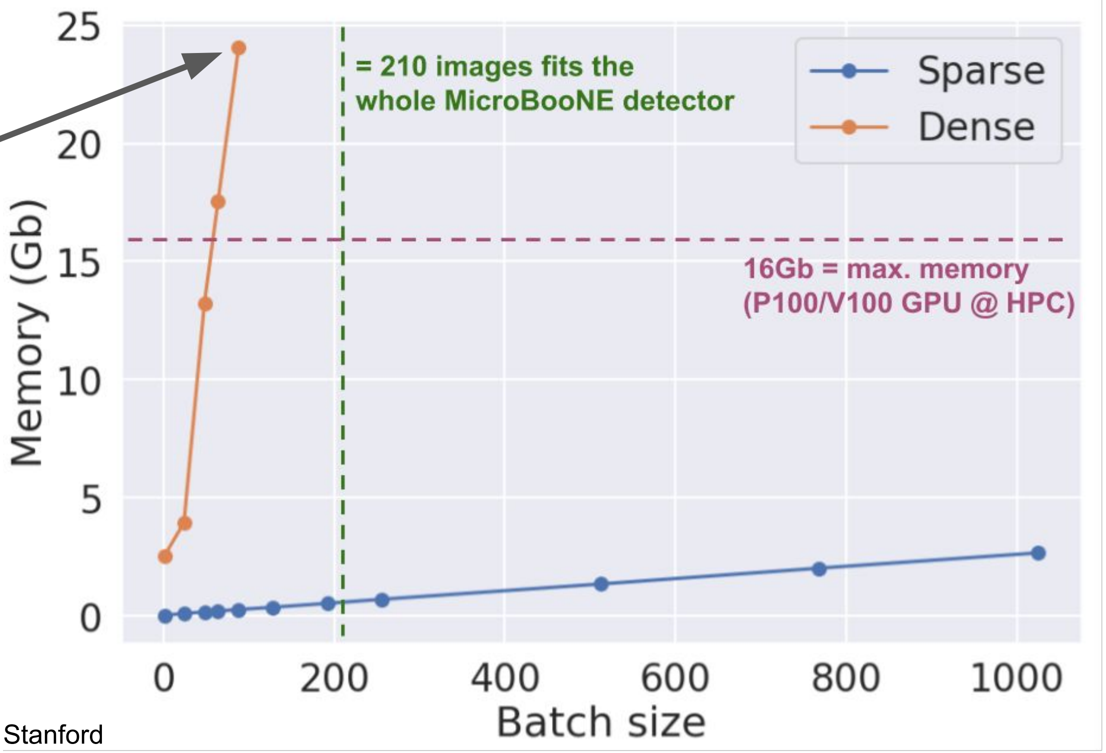


Figure credit: Laura Domine @ Stanford

Can handle easily the whole ICARUS detector which is x6 larger than MicroBooNE.

DUNE-FD is piece of cake (larger volume but less non-zero pixels)

Machine Learning & Computer Vision in Neutrino Physics

Scalable CNN for Sparse Particle Imaging Data

SLAC

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

Sparse Sub-manifold Convolutional NN

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

Computer Science > Computer Vision and Pattern Recognition

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

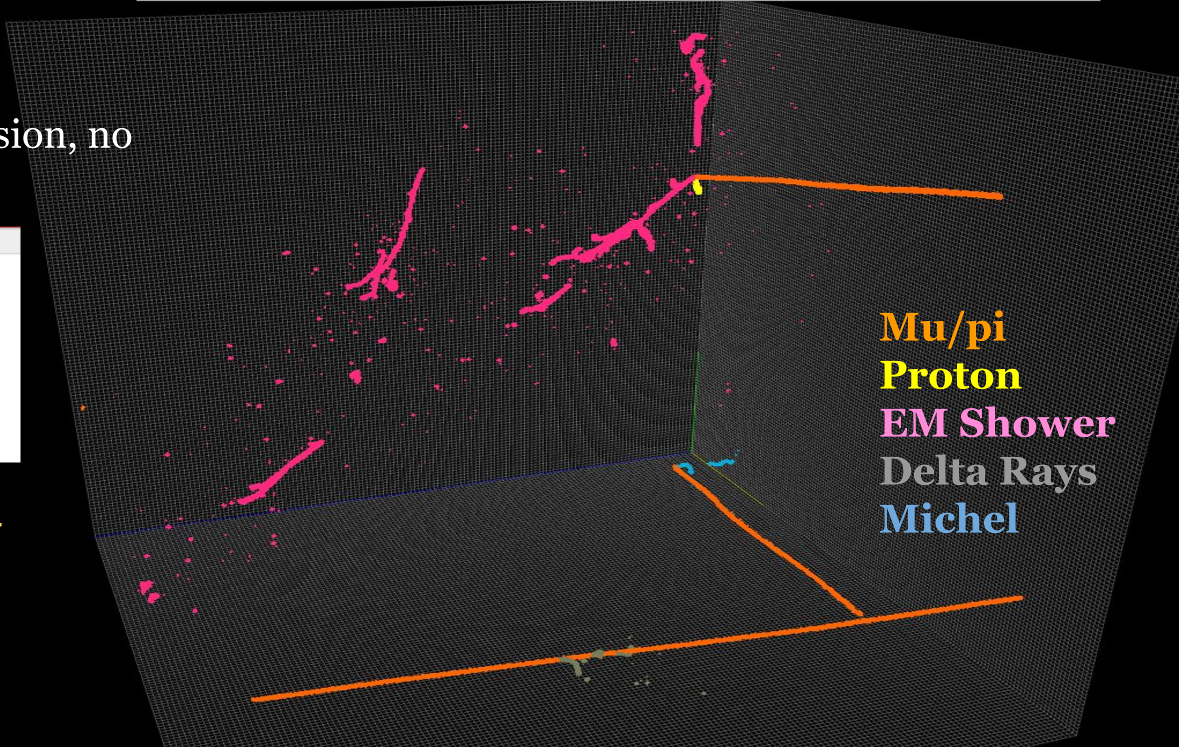
Laura Dominé, Kazuhiro Terao

(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

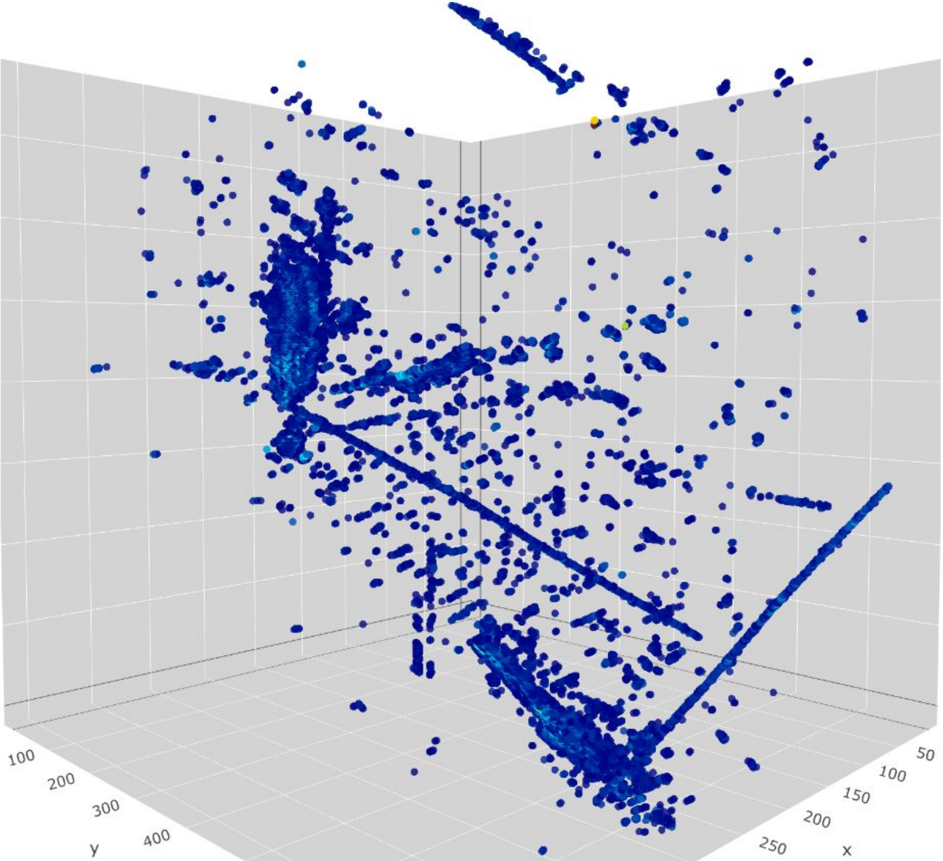
arXiv:1903.05663 presented @ ACAT 2019

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

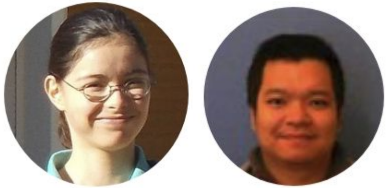


Machine Learning & Computer Vision in Neutrino Physics

Bonus: isochronous ghost point removal



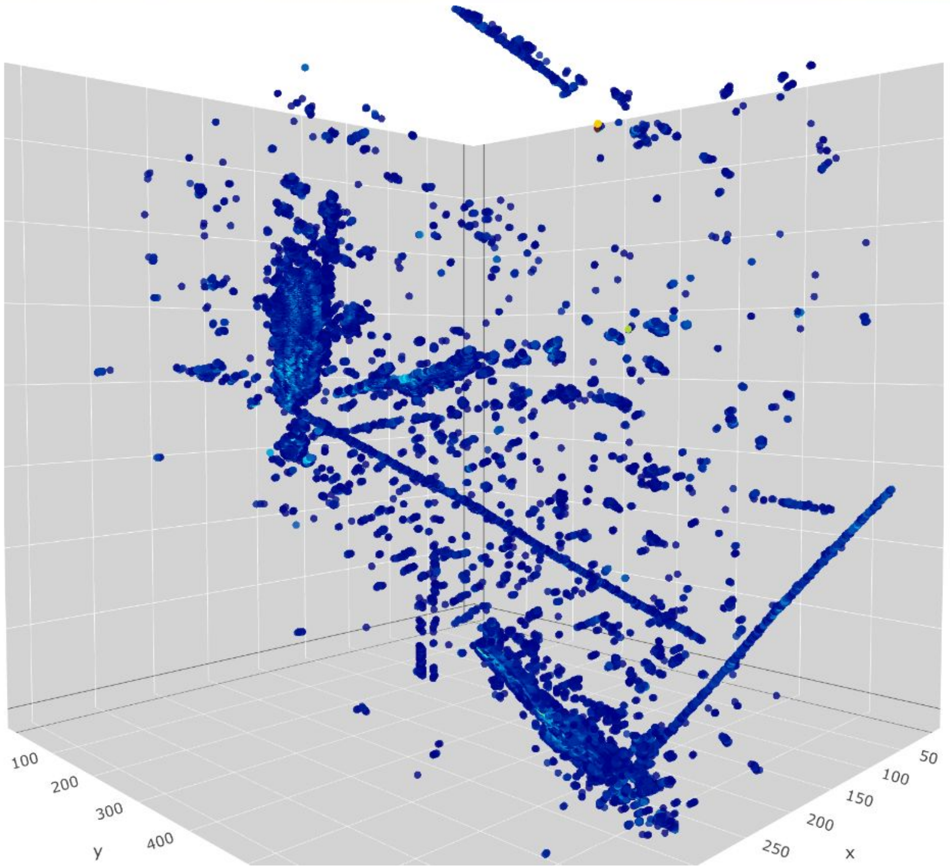
ICARUS Detector
Reconstructed 3D points

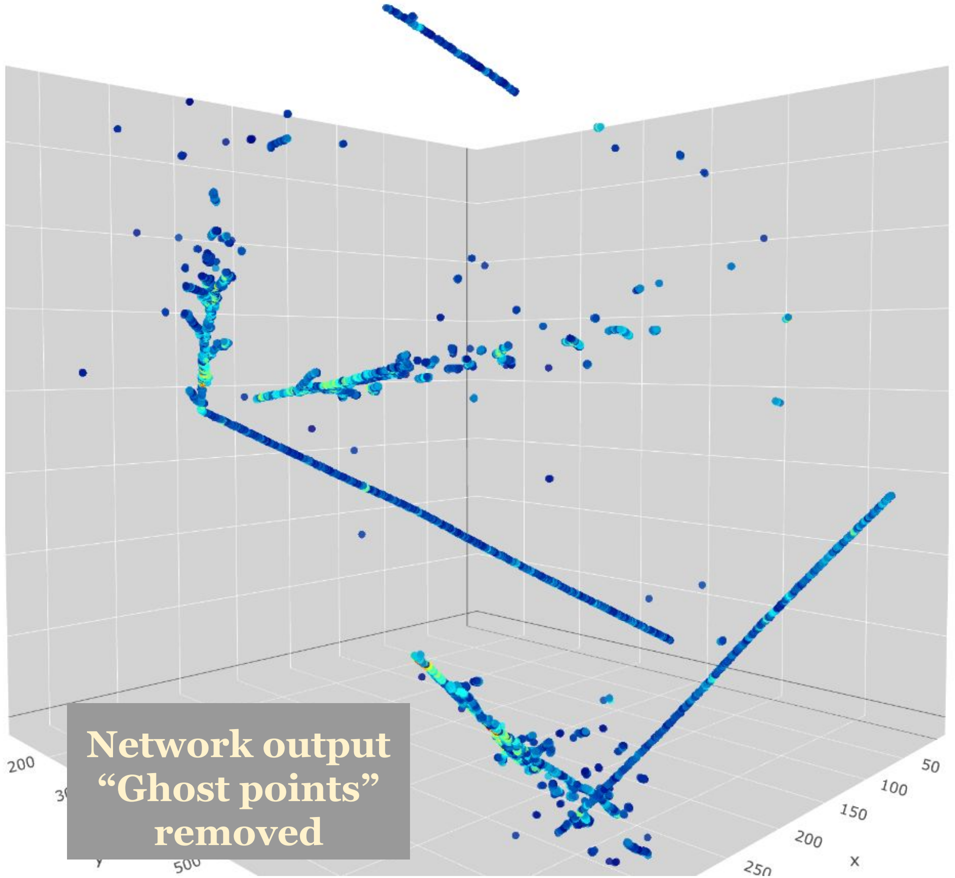


work credit:
Laura Domine
Patrick Tsang

Machine Learning & Computer Vision in Neutrino Physics

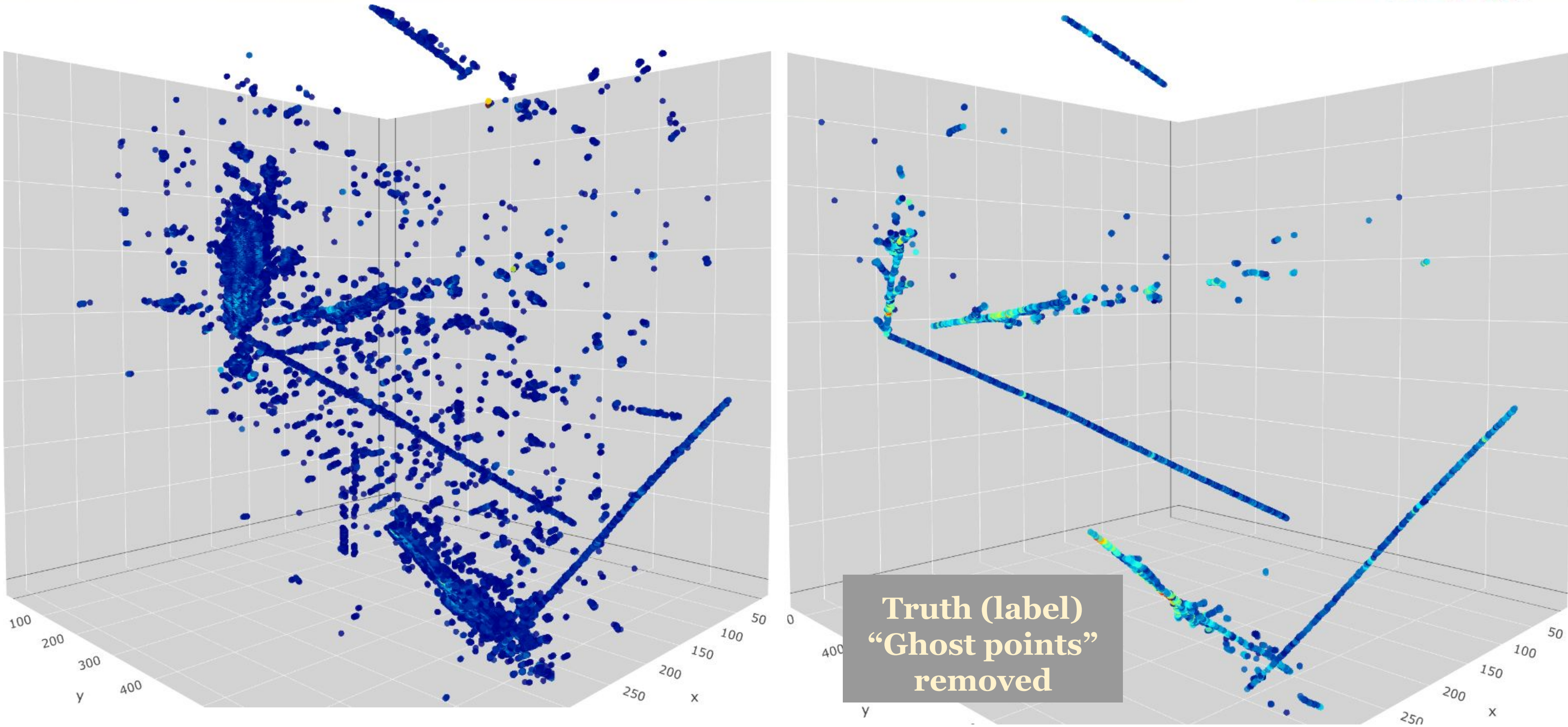
Bonus: isochronous ghost point removal





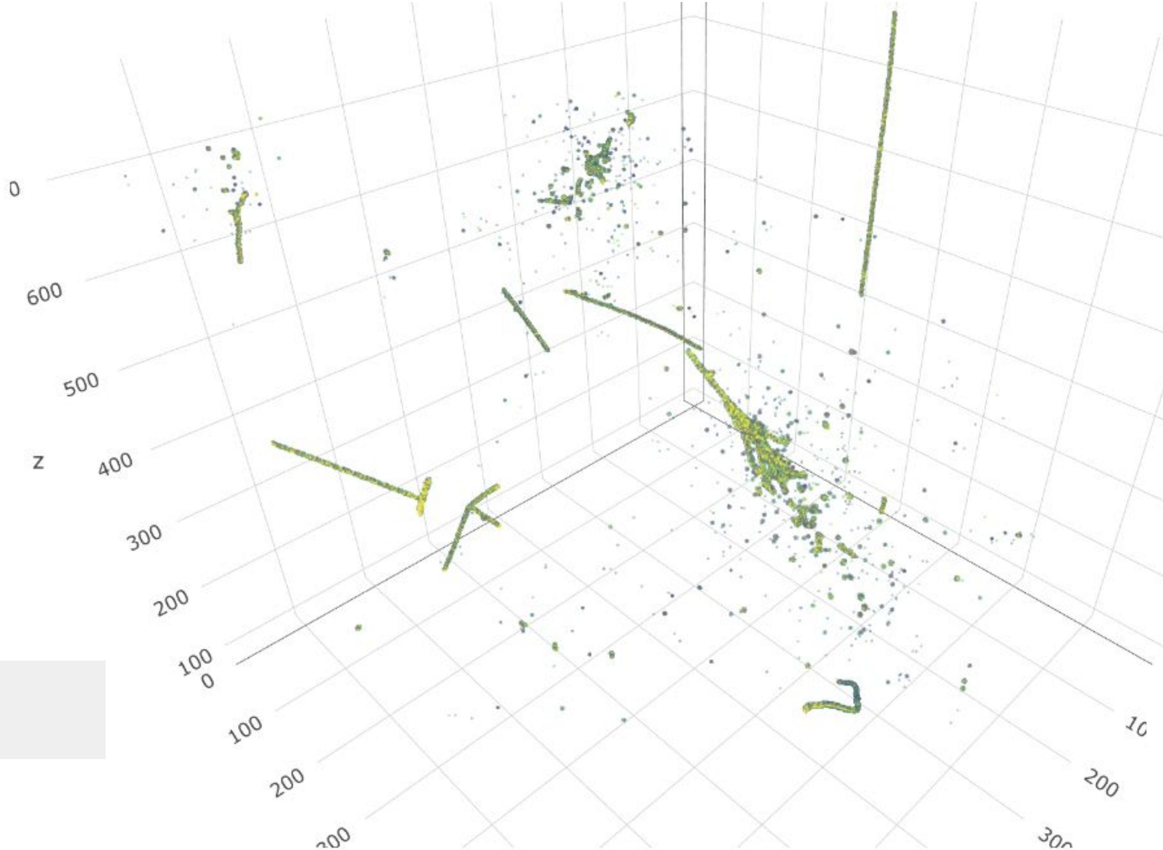
Machine Learning & Computer Vision in Neutrino Physics

Bonus: isochronous ghost point removal



Point Prediction Network (PPN)

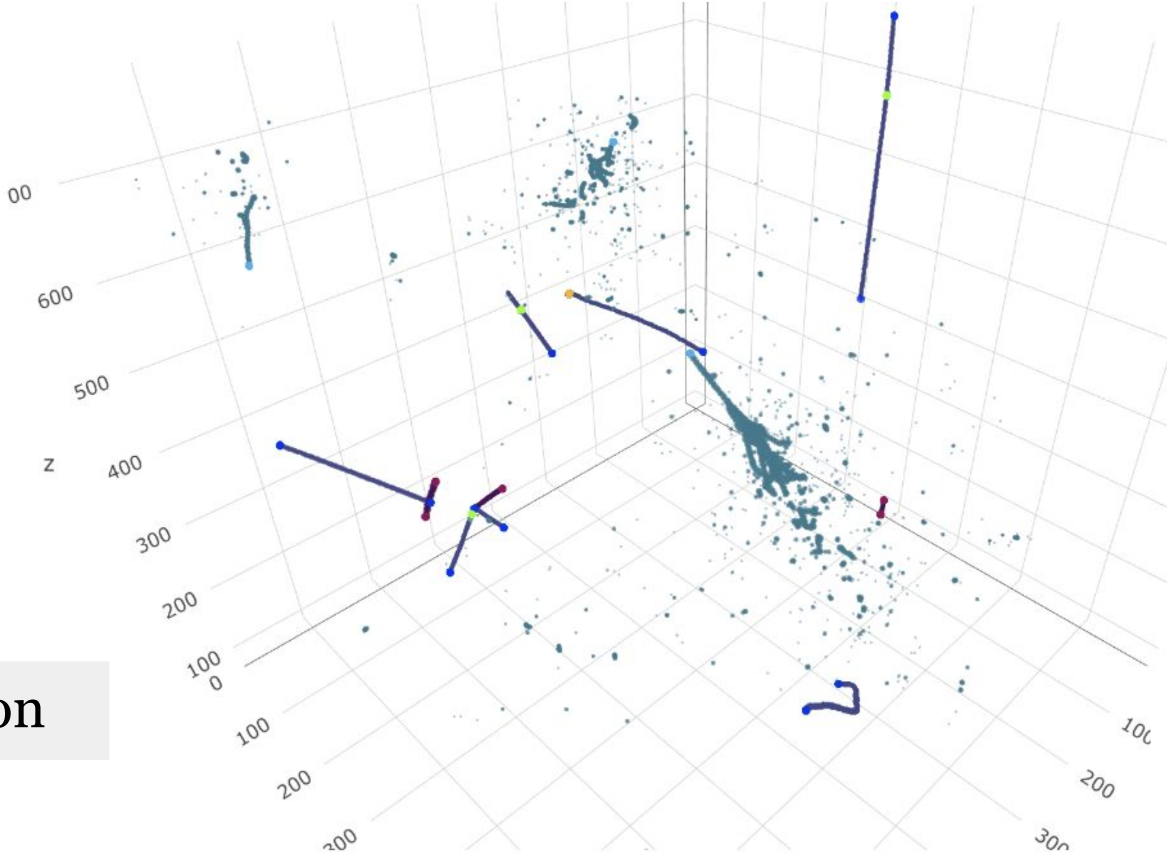
ML-based Neutrino Data Reconstruction Chain
Point Prediction Network



Input

work credit:
Laura Domine

ML-based Neutrino Data Reconstruction Chain
Point Prediction Network



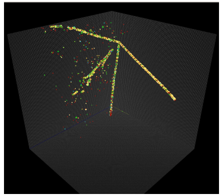
Prediction

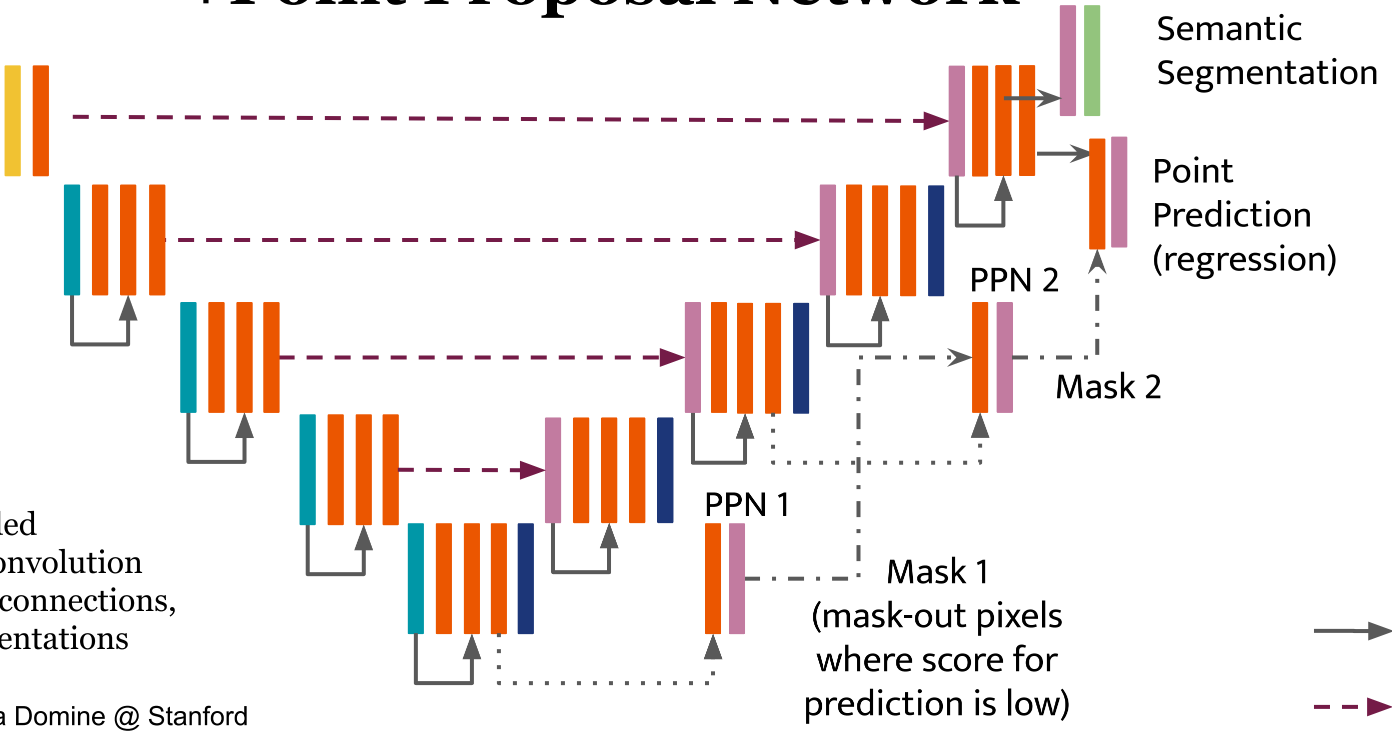
work credit:
Laura Domine

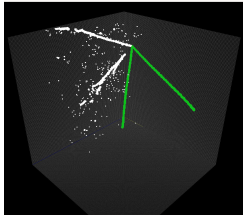
ML-based Neutrino Data Reconstruction Chain

Point Prediction Network

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



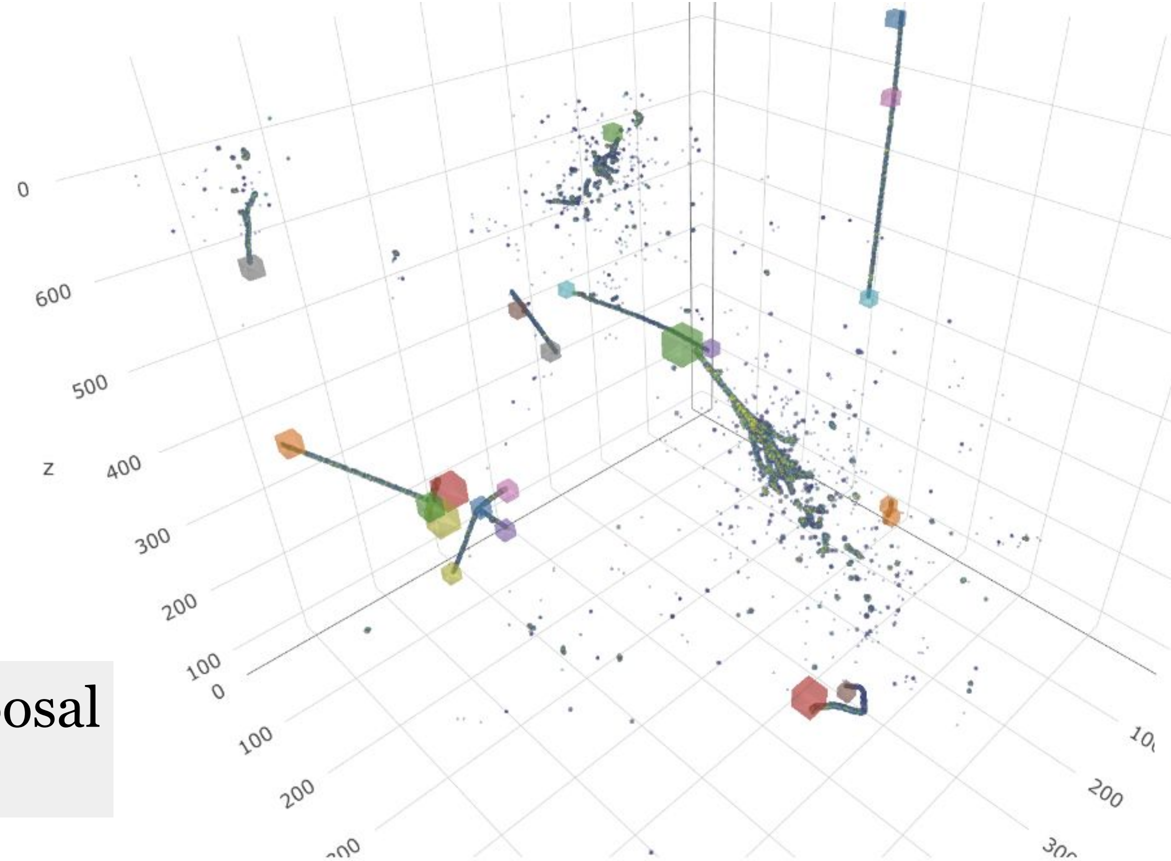




Number of strided convolutions, convolution layers, residual connections, differ in impementations

Image credit: Laura Domine @ Stanford

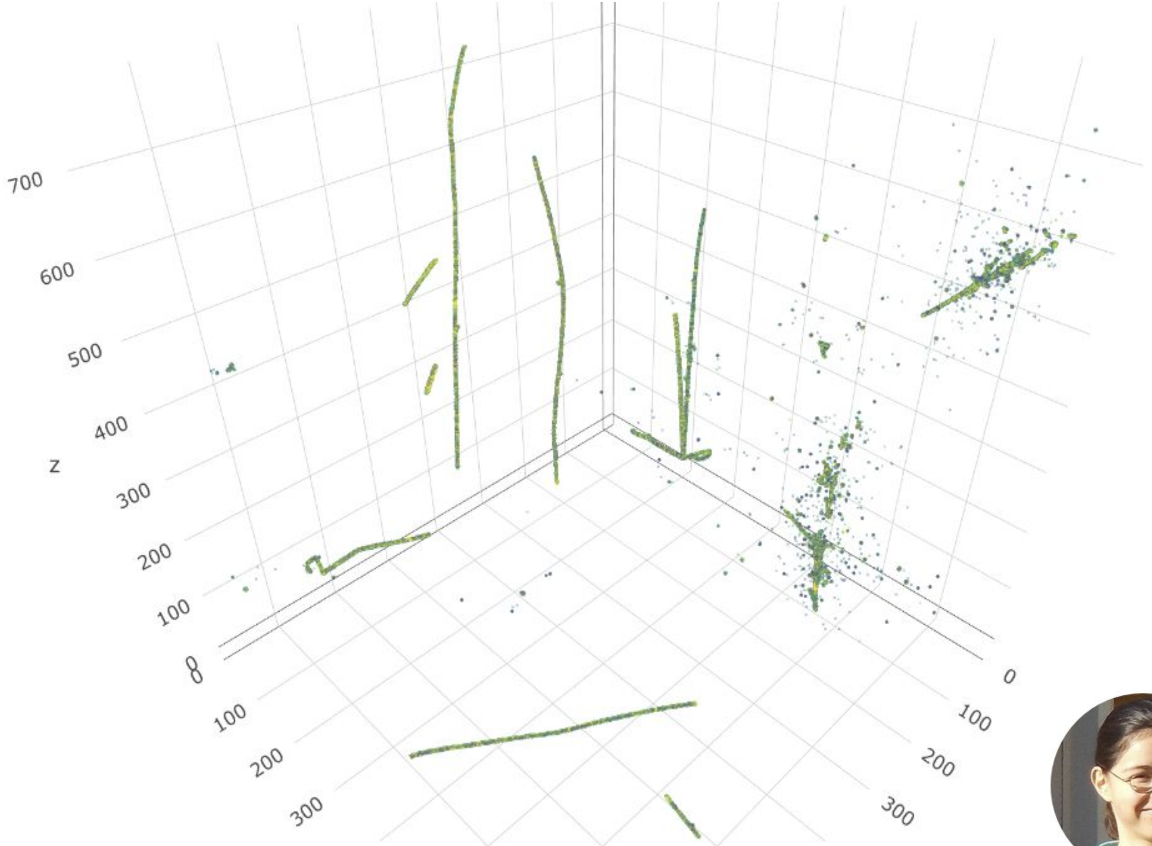
ML-based Neutrino Data Reconstruction Chain
Point Prediction Network



Deep Proposal

work credit:
Laura Domine

ML-based Neutrino Data Reconstruction Chain
Point Prediction Network

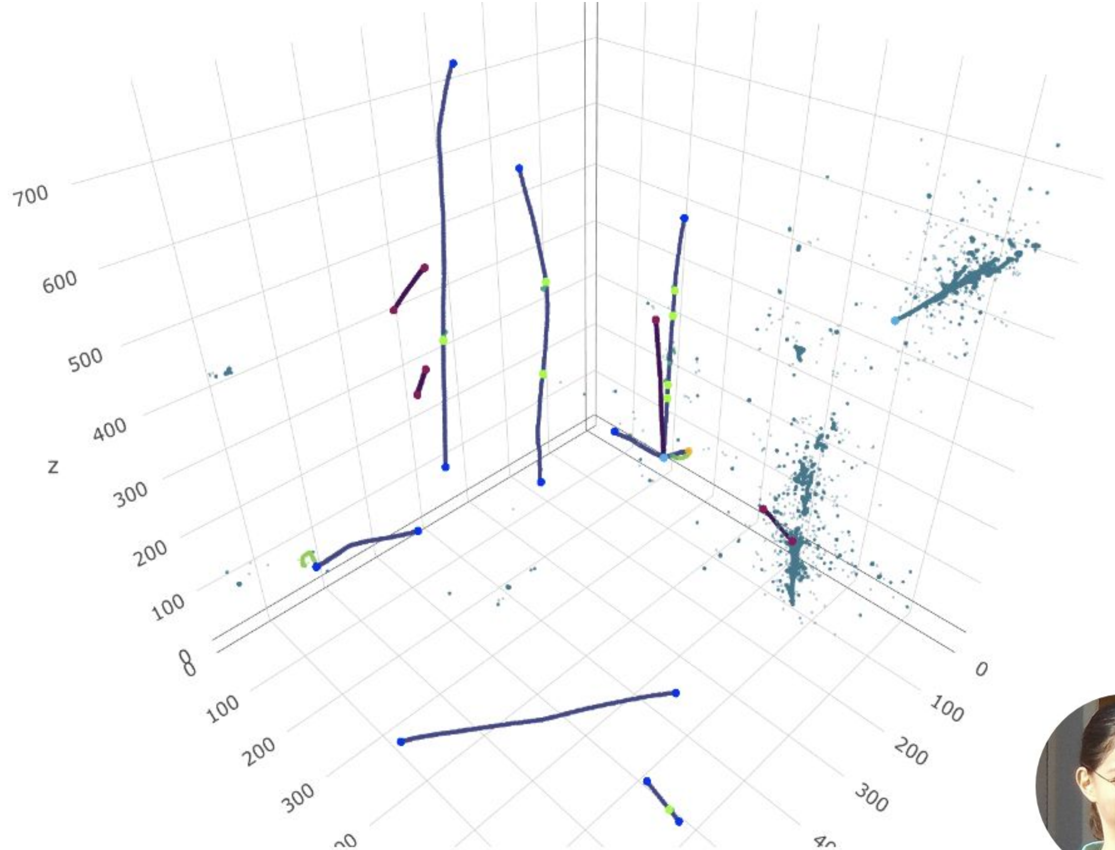


Input

work credit:
Laura Domine

ML-based Neutrino Data Reconstruction Chain

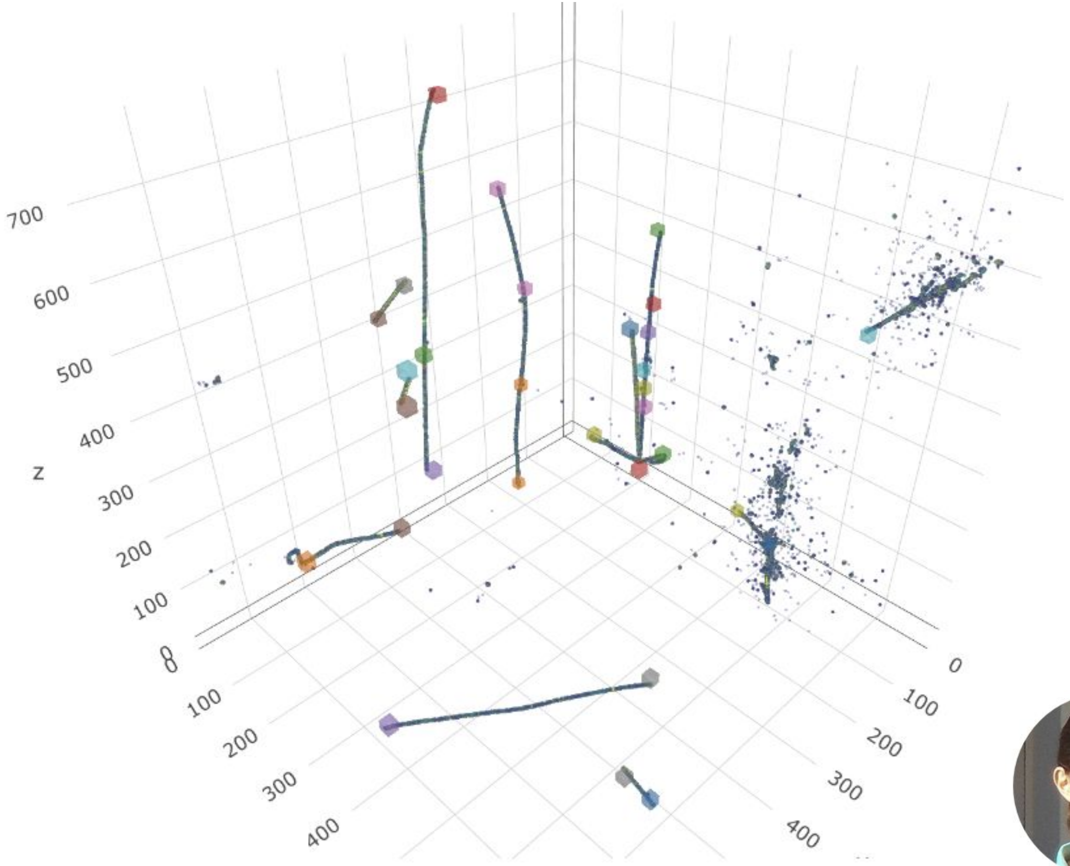
Point Prediction Network



Prediction

work credit:
Laura Domine

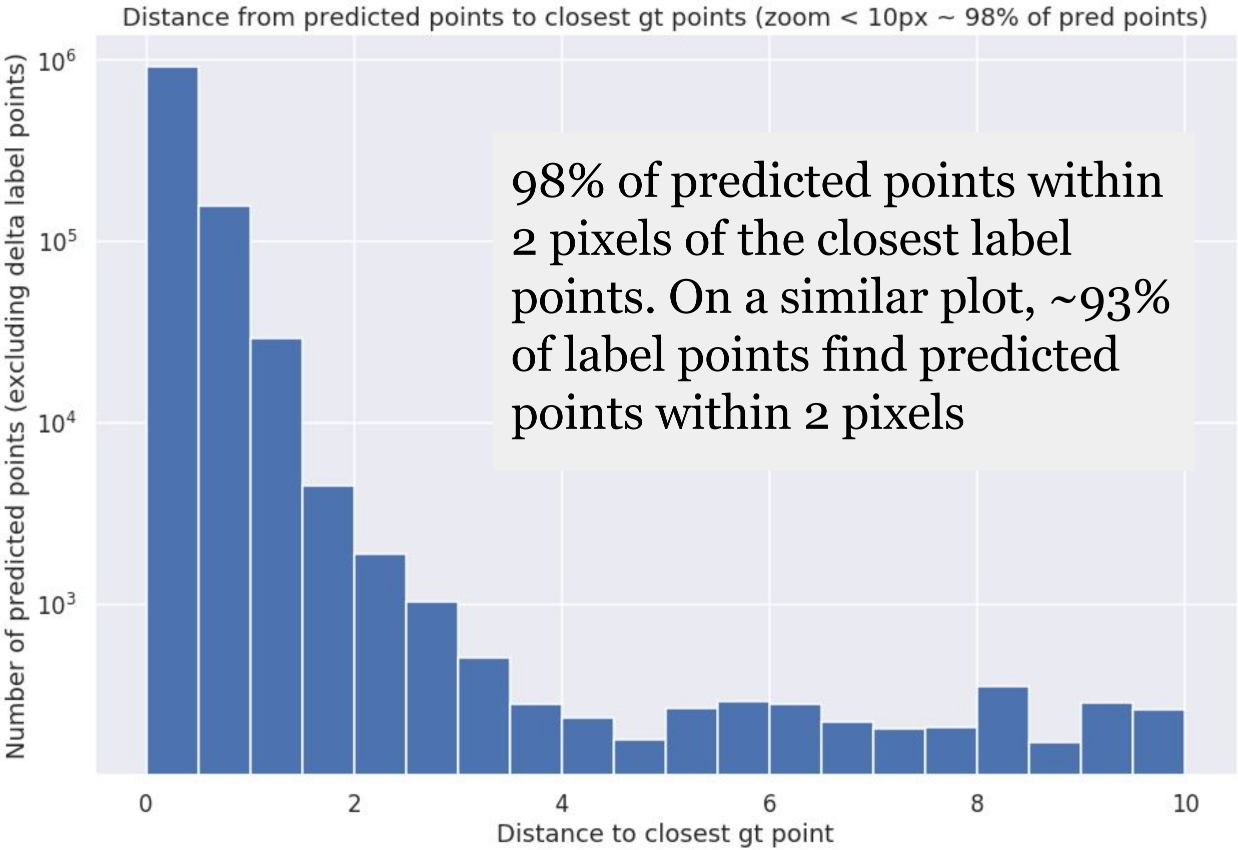
ML-based Neutrino Data Reconstruction Chain
Point Prediction Network

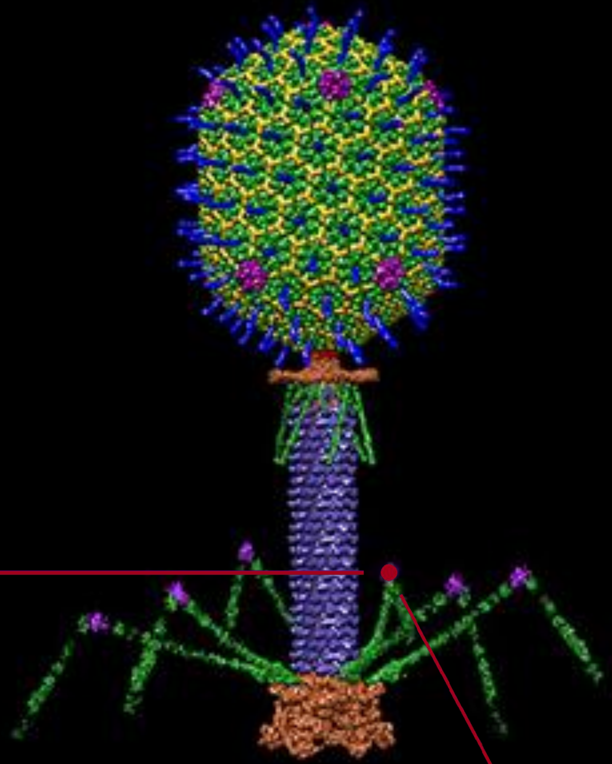


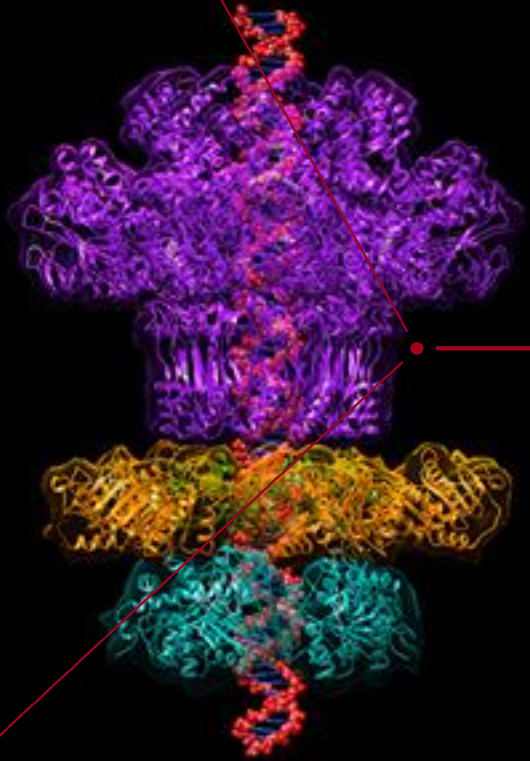
Deep Proposal

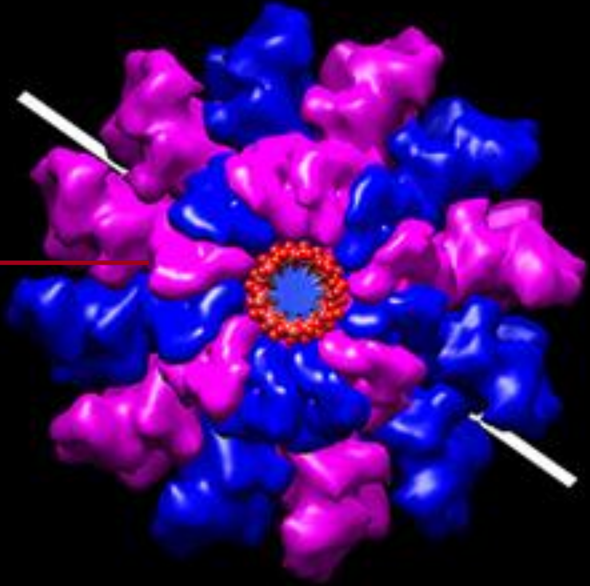
work credit:
Laura Domine

ML-based Neutrino Data Reconstruction Chain
Point Prediction Network





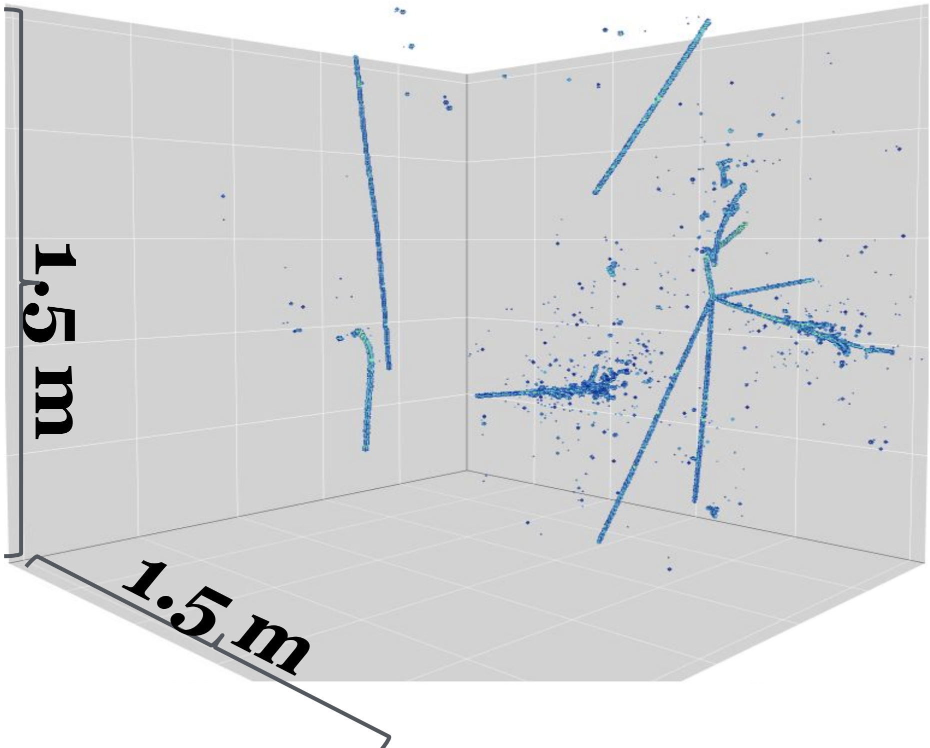




Clustering

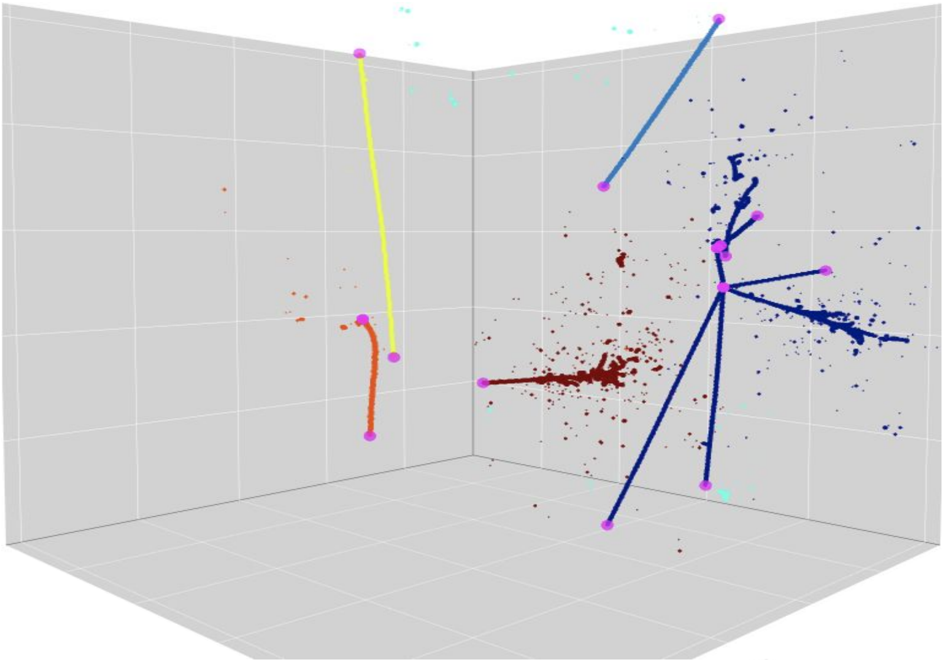
Particle & Interaction Clustering

Goal: group pixels into interesting unit of instance

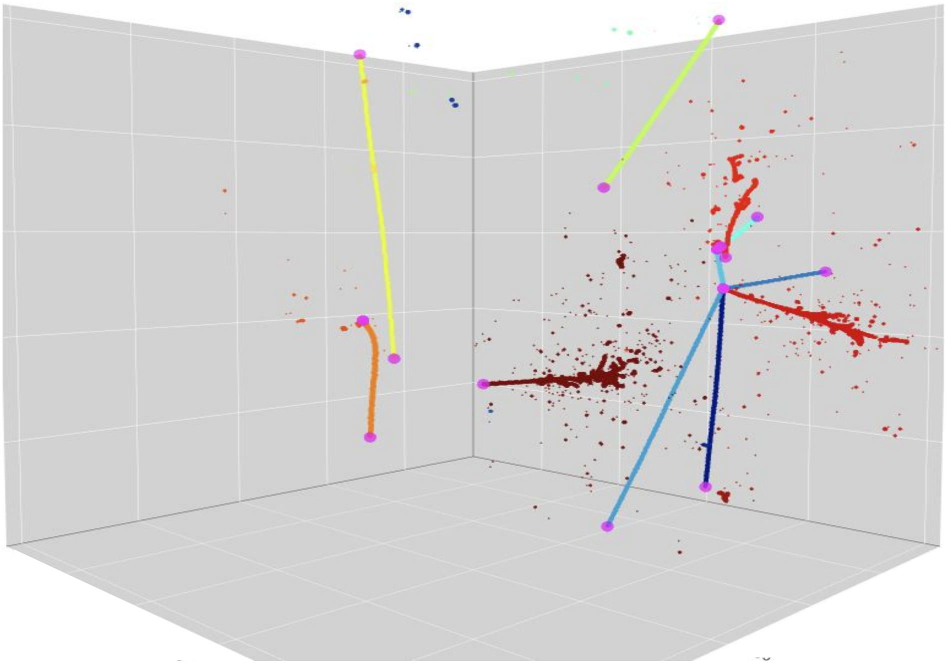


ML-based Neutrino Data Reconstruction Chain
Particle & Interaction Clustering

Goal: group pixels into interesting unit of instance



Interaction



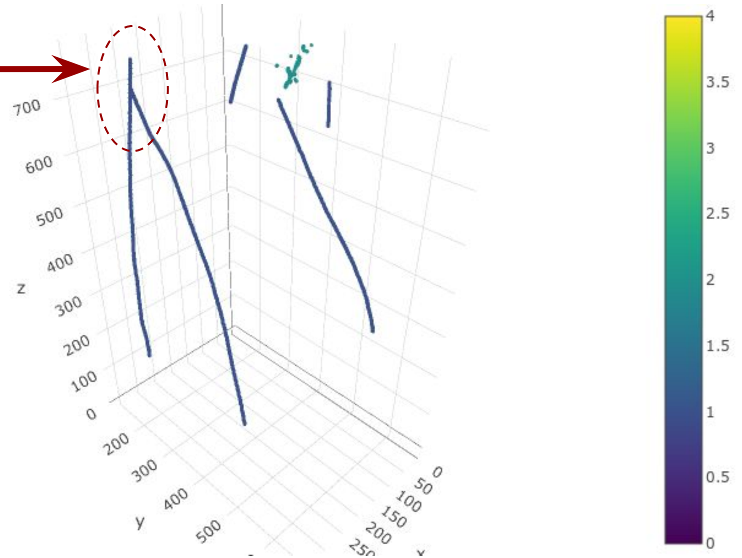
Particle

Particle & Interaction Clustering

Wait… how about just DBSCAN (or spectral clustering, etc.) in 3D?

- Would be great! but like other non-ML approach, need many heuristics at the end + won't do well enough

Example: two co-linear particles with the same start point (the right particle branched off due to scattering), which is an important physics feature (i.e. we don't want to exclude).
… DBSCAN connects two tracks together.

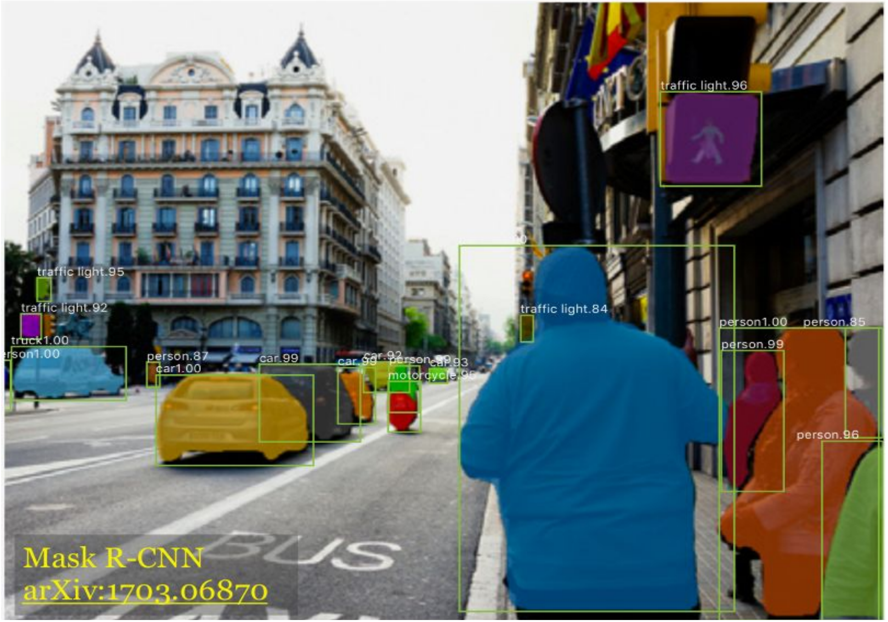


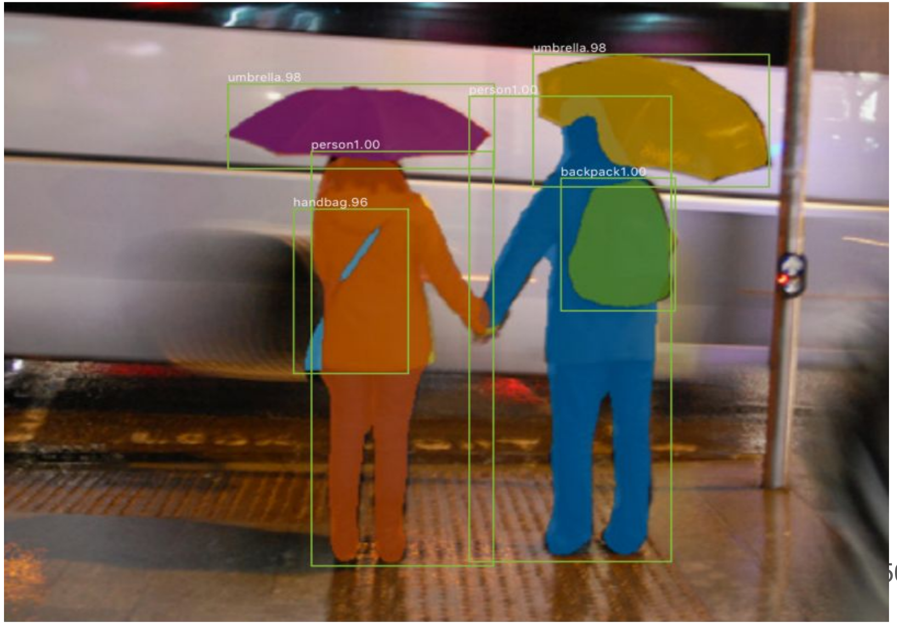
ML-based Neutrino Data Reconstruction Chain
Particle & Interaction Clustering

SLAC

Instance+Semantic Segmentation

- **Mask R-CNN** … a popular solution, many applications in science/industries
  - Object (=instance) detection + instance pixel masking in a bounding box

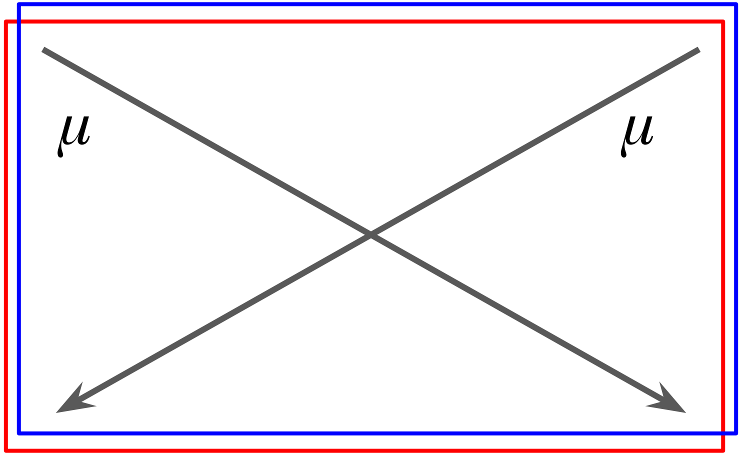


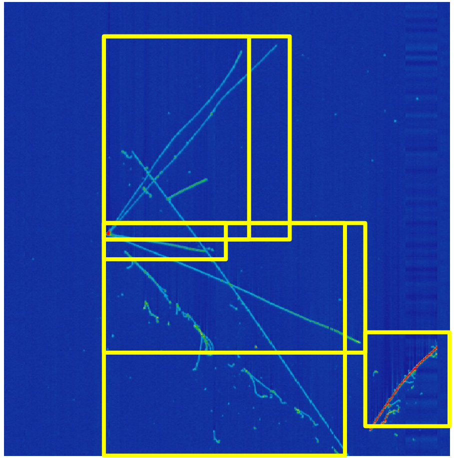


Particle & Interaction Clustering

Instance+Semantic Segmentation

- **Mask R-CNN** … a popular solution, many applications in science/industries
  - Object (=instance) detection + instance pixel masking in a bounding box
  - **Challenge**: instance distinction is affected by BB position/size
  - Another family: Single-Shot-Detection (SSD) based (not covered here)





Occlusion issue

The overlap rate of particles can be challenging!

Particle & Interaction Clustering

Instance+Semantic Segmentation

- **Alternative**: transform data into easily clusterable hyperspace
  - Network learns how to transform from the feature space to hyperspace

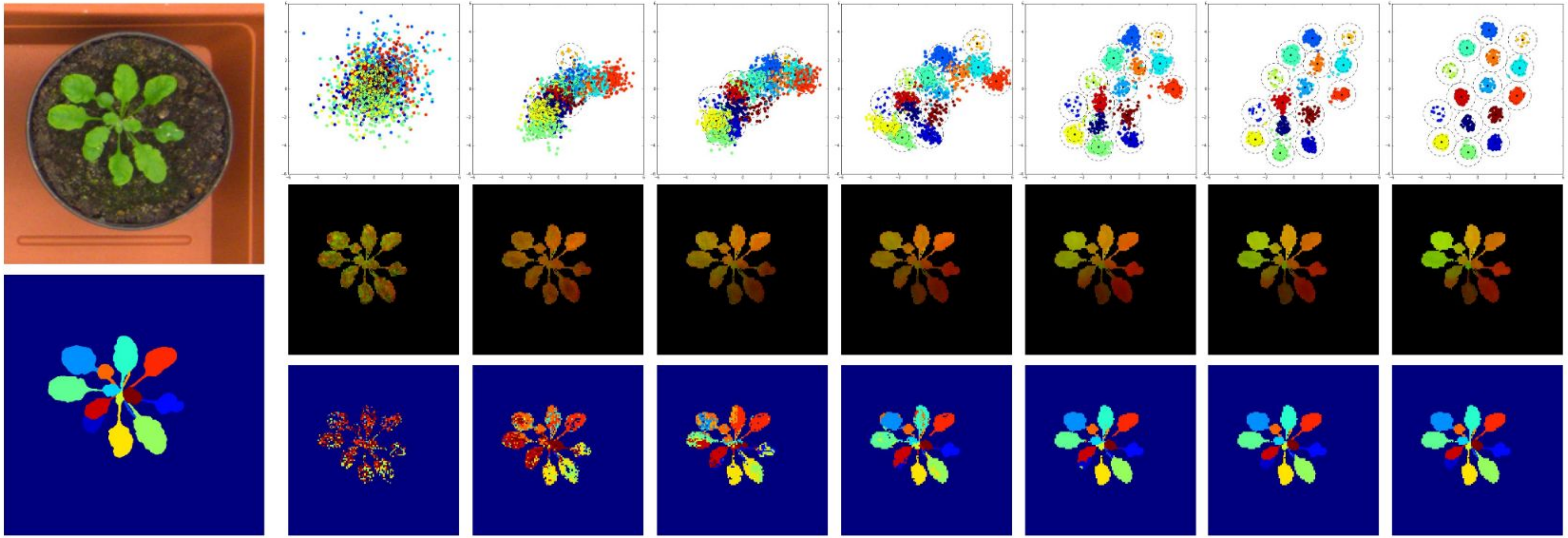
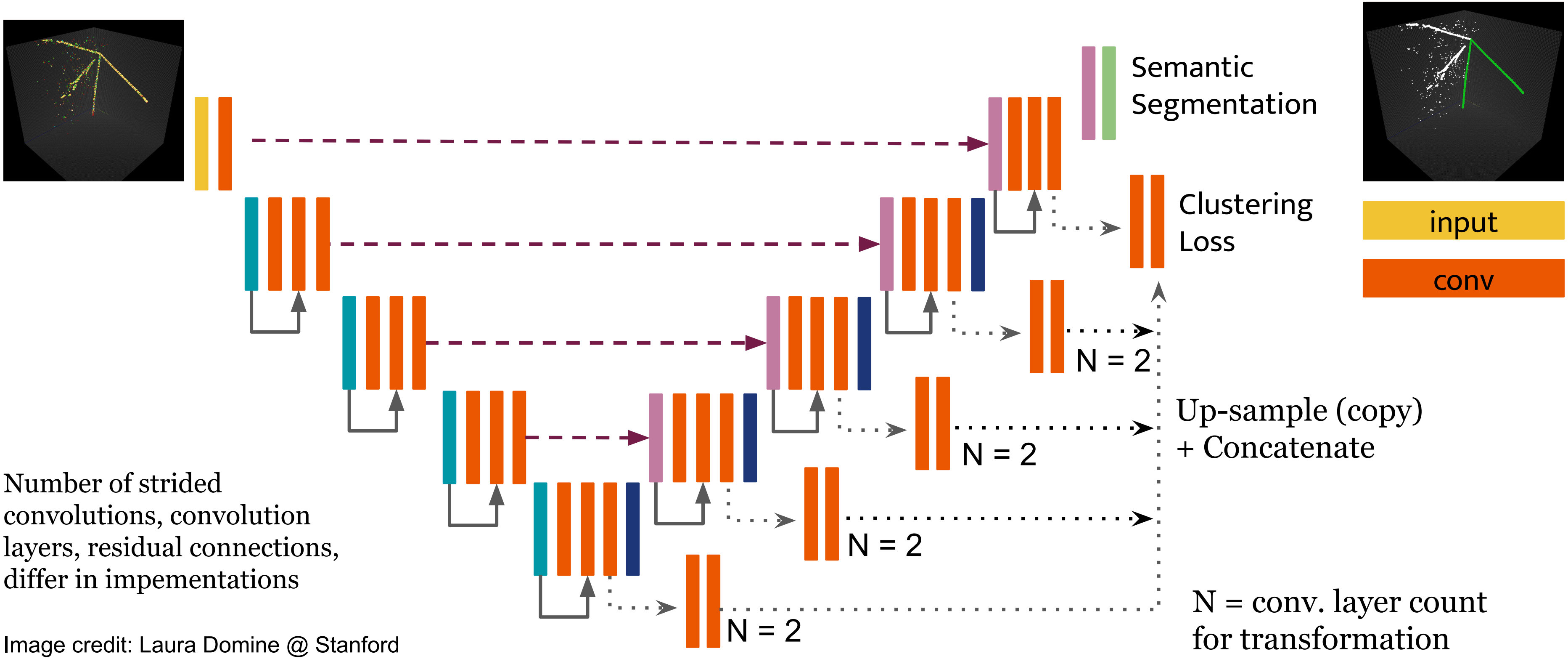


Image credit: arXiv 1708.02551

ML-based Neutrino Data Reconstruction Chain
Particle & Interaction Clustering

Instance+Semantic Segmentation



Number of strided convolutions, convolution layers, residual connections, differ in impementations

Image credit: Laura Domine @ Stanford

Particle & Interaction Clustering

Instance+Semantic Segmentation

- Three component loss: pull together points that belong to the same cluster, keep distance between clusters, and regularization

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

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

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

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

Equation credit: Dae Hyun K. @ Stanford

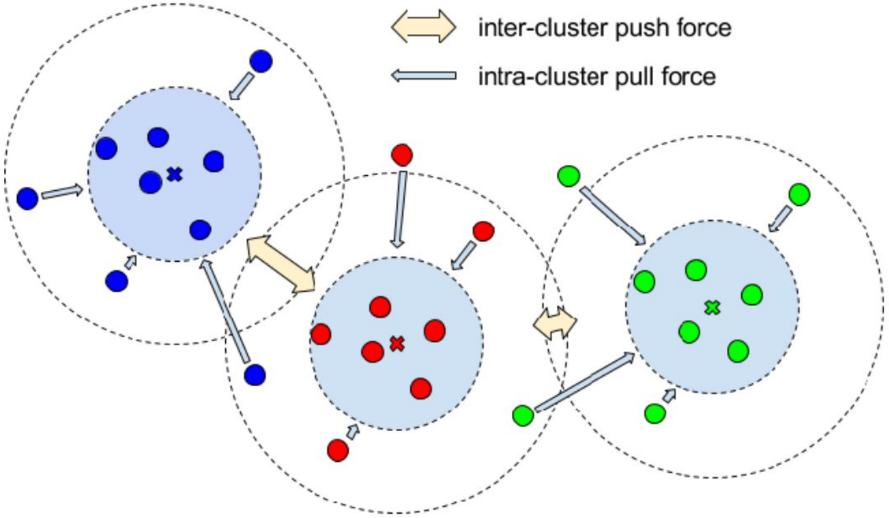
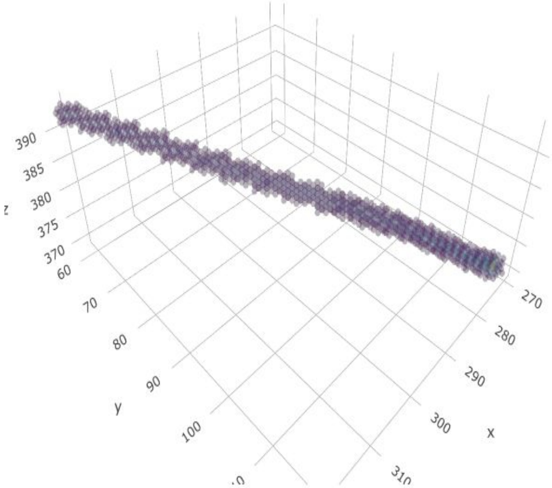


Image credit: arXiv 1708.02551

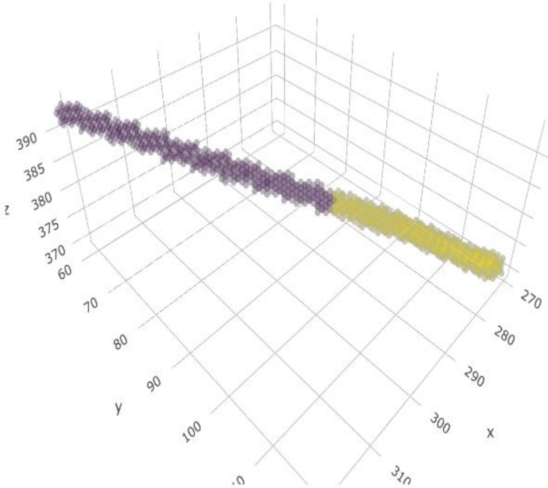
ML-based Neutrino Data Reconstruction Chain
Particle & Interaction Clustering

Instance+Semantic Segmentation

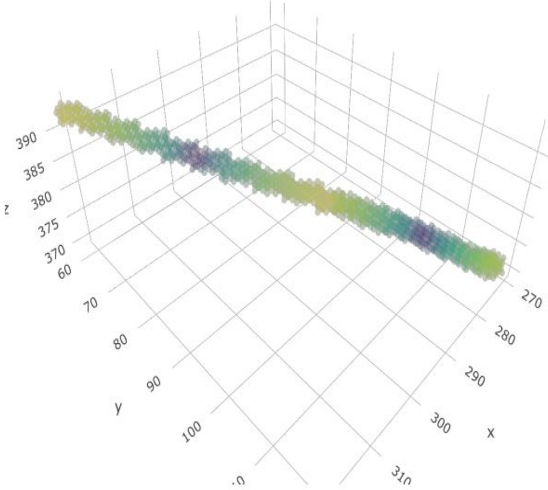
- The distance from the cluster center to each point in the hyper-space (below right), indicating radial point distributions.



Energy Deposition



Clustered



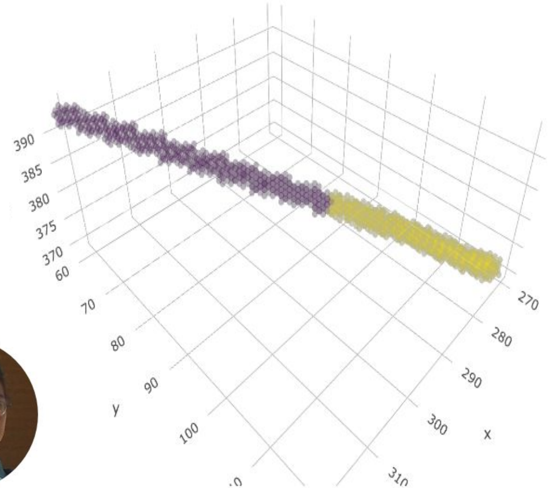
Hyper-space Dist.

Particle & Interaction Clustering

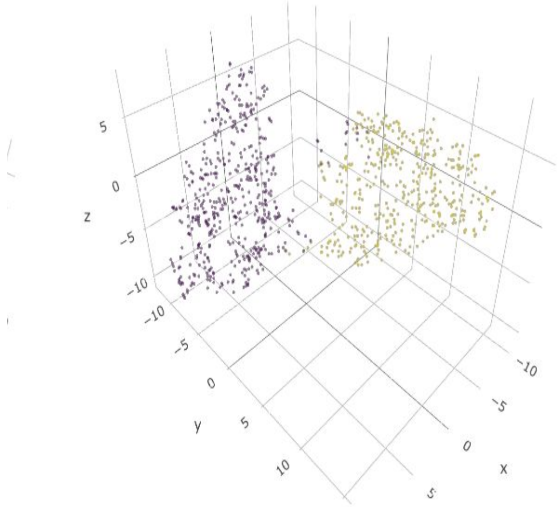
Instance+Semantic Segmentation

- The distance from the cluster center to each point in the hyper-space (below right), indicating radial point distributions.

work credit:
Dae Heun Koh
Laura Domine



Clustered

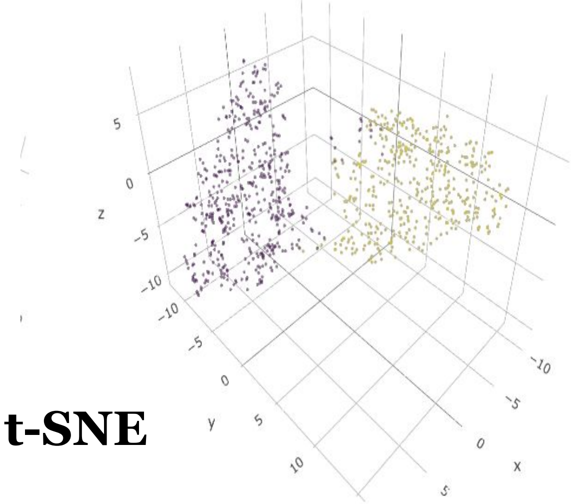


t-SNE

Clustering in hyperspace?

A. Clustering-in-the-network (not covered here, ask me later!)
B. Optimize network for post-processing clustering (e.g. DBSCAN)

Example: For B) + DBSCAN, how to optimize?



t-SNE

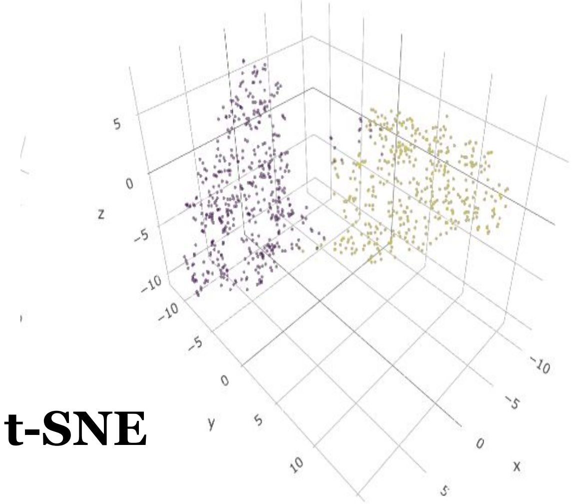
Ultimately, DBSCAN clusters neighbor points within dist. ϵ...

- **+ Two "point density" loss**
  - Maximize density of points that belong to the same cluster
  - Minimize density of points that belong to a different cluster

Clustering in hyperspace?

A. Clustering-in-the-network (not covered here, ask me later!)
B. Optimize network for post-processing clustering (e.g. DBSCAN)

Example: For B) + DBSCAN, how to optimize?



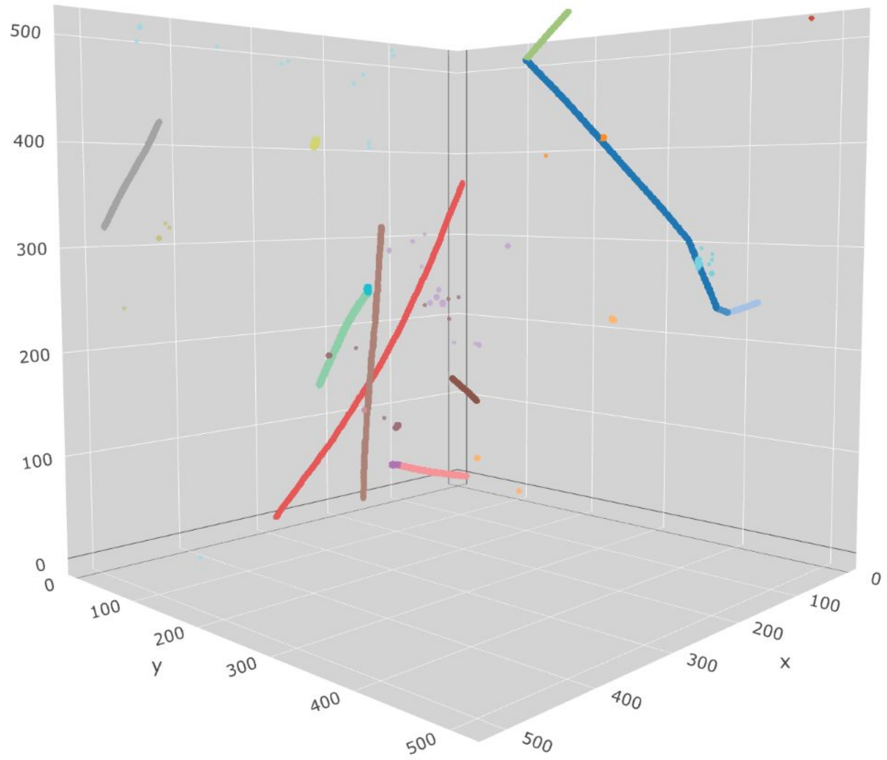
t-SNE

Ultimately, DBSCAN clusters neighbor points within dist. ϵ…

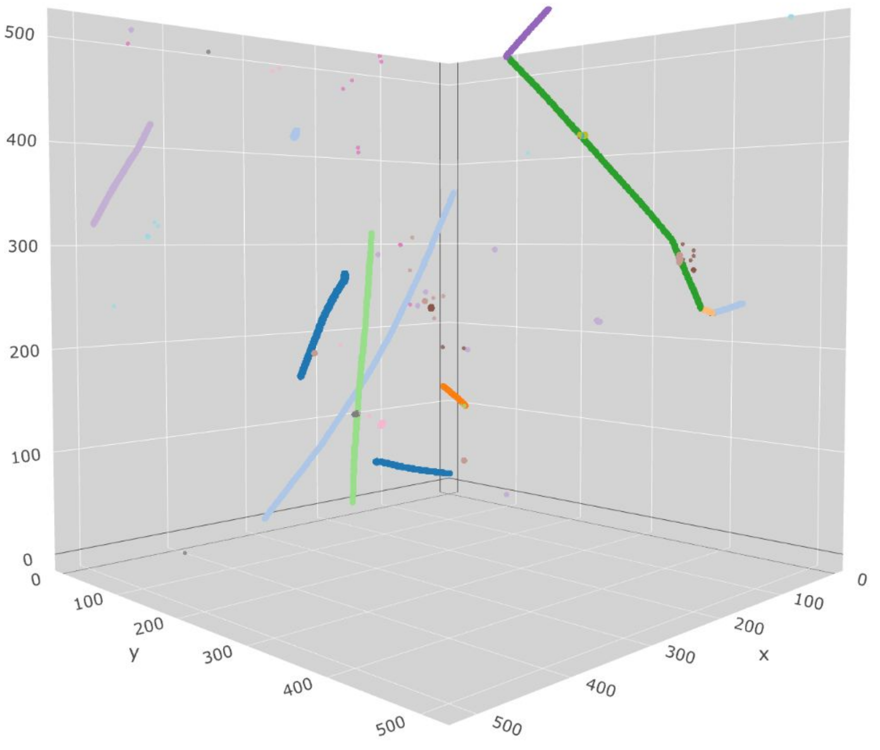
- **+ Two "point density" loss**
  - Maximize density of points that belong to the same cluster
  - Minimize density of points that belong to a different cluster
- **+ Two "point density estimate" target + loss**
  - Estimate same vs. different cluster point densities (2 floats)
  - These estimates serve to identify "boundary points".
  - At inference time, run DBSCAN without boundary points first to separate clusters.

ML-based Neutrino Data Reconstruction Chain
Particle & Interaction Clustering

Truth (label)



Prediction



Play with this and that examples!

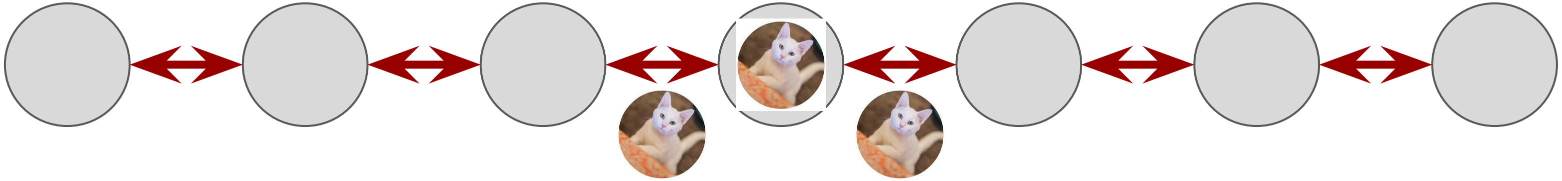
These are randomly selected events where many showers (first example) and many tracks (second example) are found…

Graph Neural Network

- Based on "*nodes*" and "*edges*" … feature propagates by "message passing" (MP)

Graph Neural Network

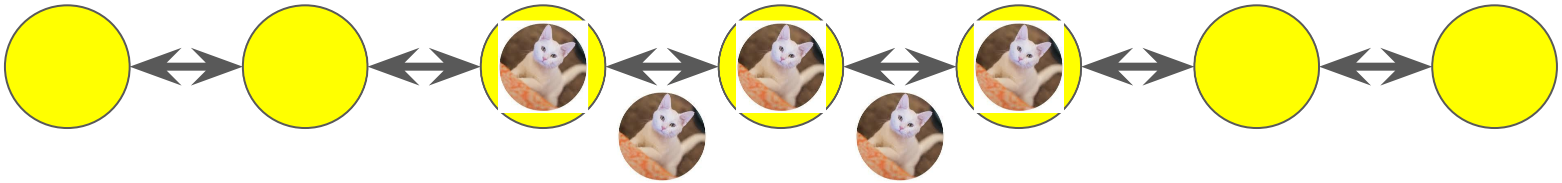
- Based on "*nodes*" and "*edges*" ... feature propagates by "message passing" (MP)



ML-based Neutrino Data Reconstruction Chain
Particle & Interaction Clustering

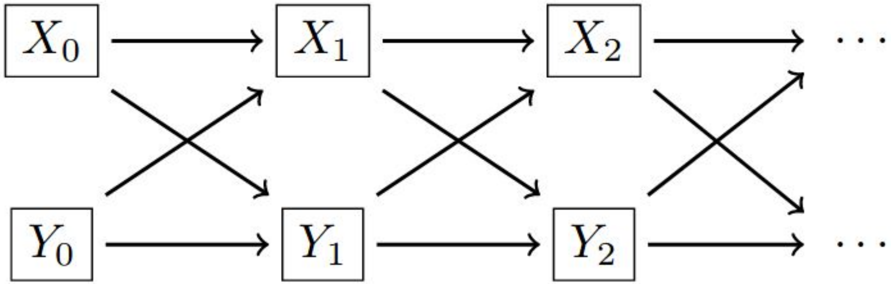
Graph Neural Network

- Based on "*nodes*" and "*edges*" ... feature propagates by "message passing" (MP)



Quick GNN overview

- 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)

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

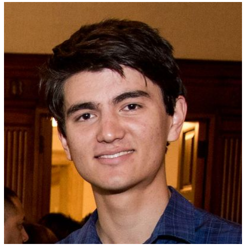
- Node feature at i, layer k+1

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

Graph Neural Network

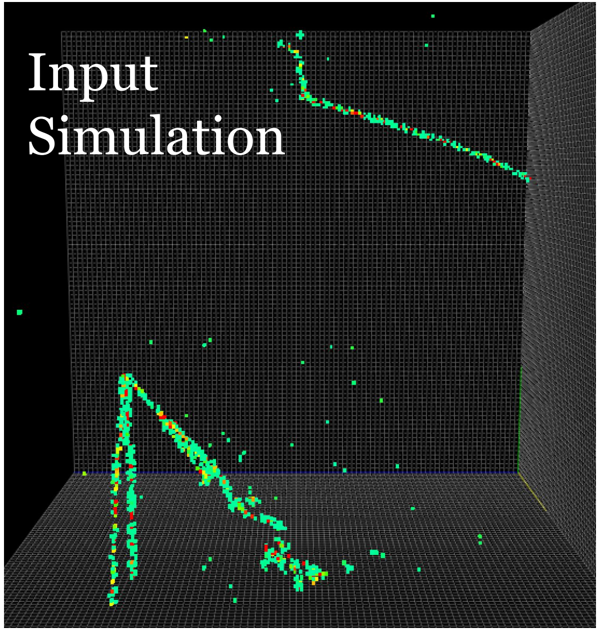
- Based on "*nodes*" and "*edges*" … feature propagates by "message passing" (MP)
- Type of GNNs:
  - Graph structure (e.g. fully-connected vs. bipartite graph)
  - Message passing (e.g. MLP, attention, LSTM, etc.)
  - Dynamic vs. Static graph
    - Example A: semantic segmentation using dynamic graph
    - Example B: Shower hierarchy using static (bipartite) graph

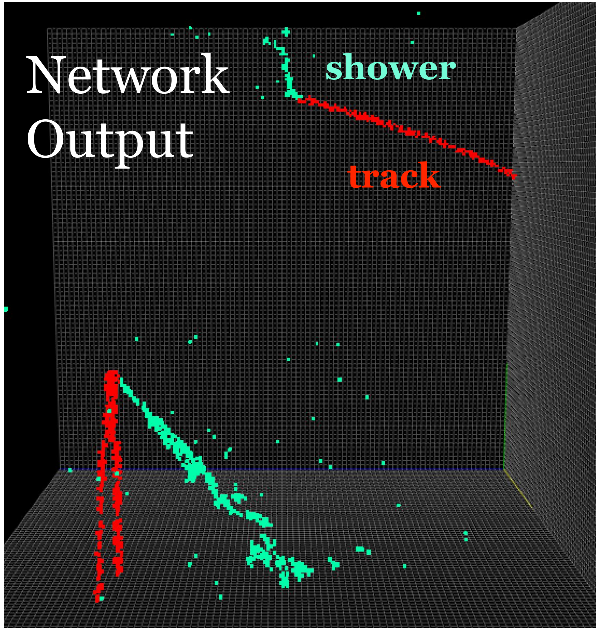




work credit:
Dae Heun Koh
Laura Domine
Alexander Zlokapa

Semantic Segmentation w/ Dynamic GNN

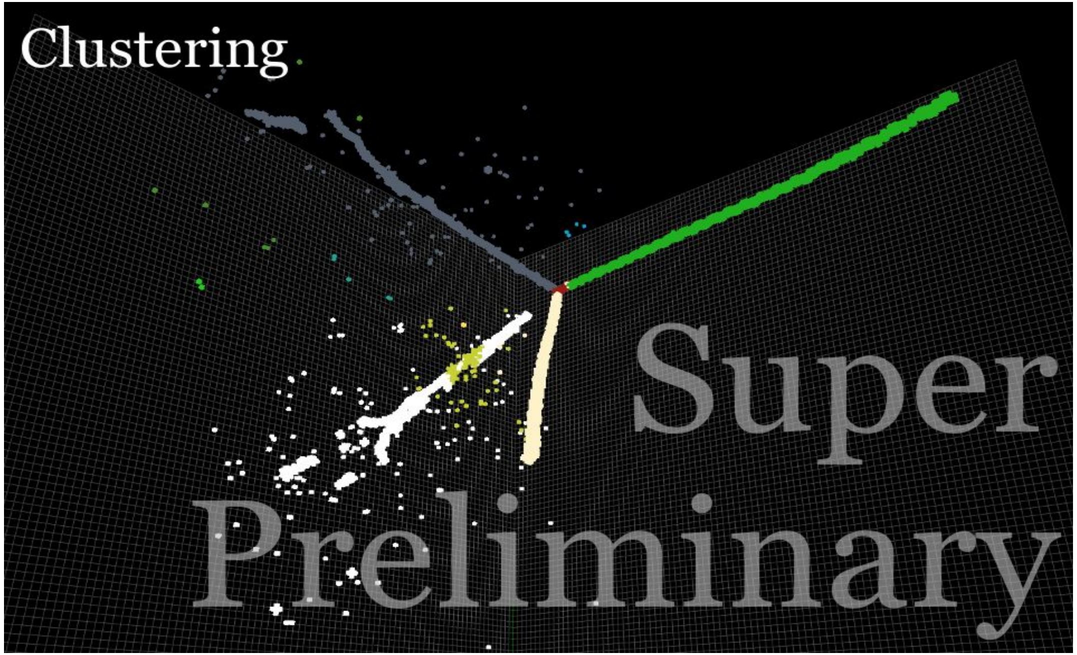




- Node = every 3D voxel
- MP with 40 "neighbors"
- Edge re-defined after 2 convolutions, repeat for 3
- Left: sparsified LArTPC trajectories from R&D work, accuracy ~97% (compared to >99% of SCN)

CNN is really good for densely connected pixel features. GNN's strength (distant feature correlation) not in effect here.

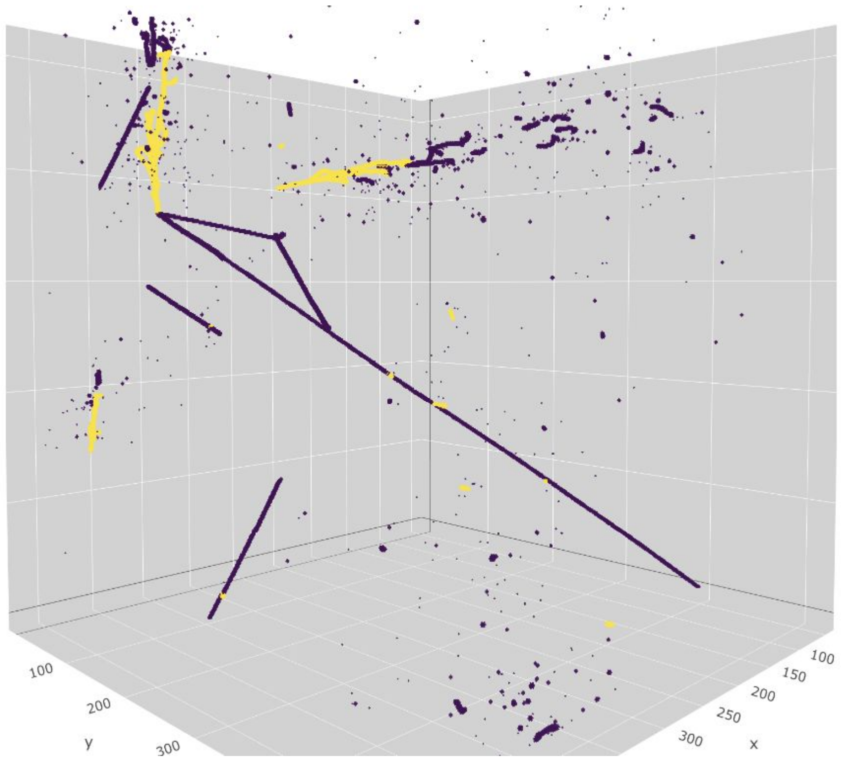
Clustering w/ Dynamic GNN

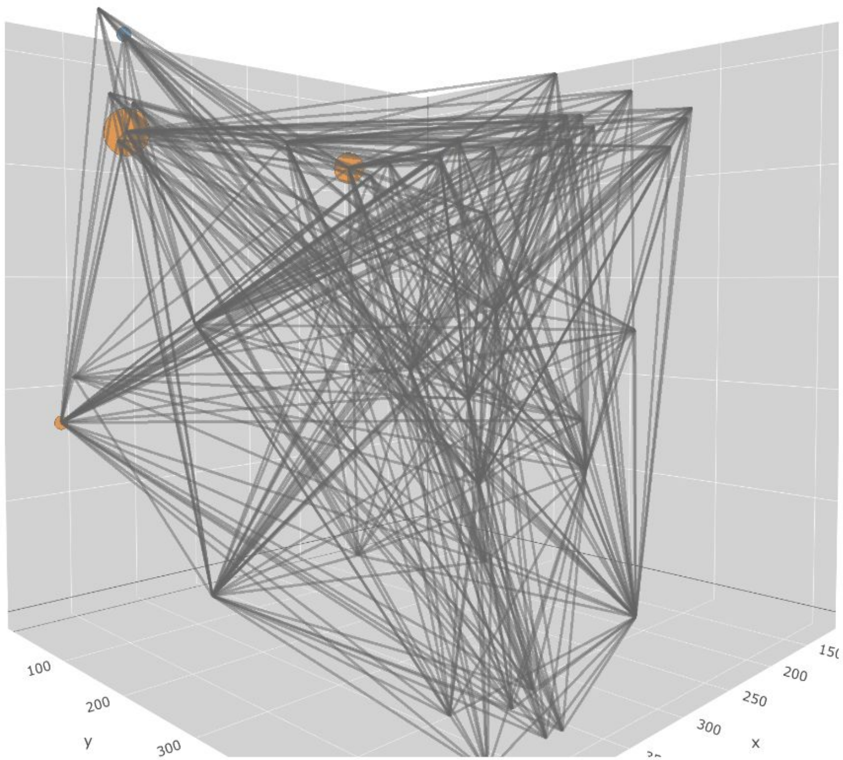


- Node = every 3D voxel
- MP with 40 “neighbors”
- Edge re-defined after 2 convolutions, repeat for 3
- Left: small shower mis-clustered
- **Issue**: to work well, cannot be sparsified + much harder to learn the same level performance compared to CNN **because** it has an extra task of learning “optimal graph structure (edges)”

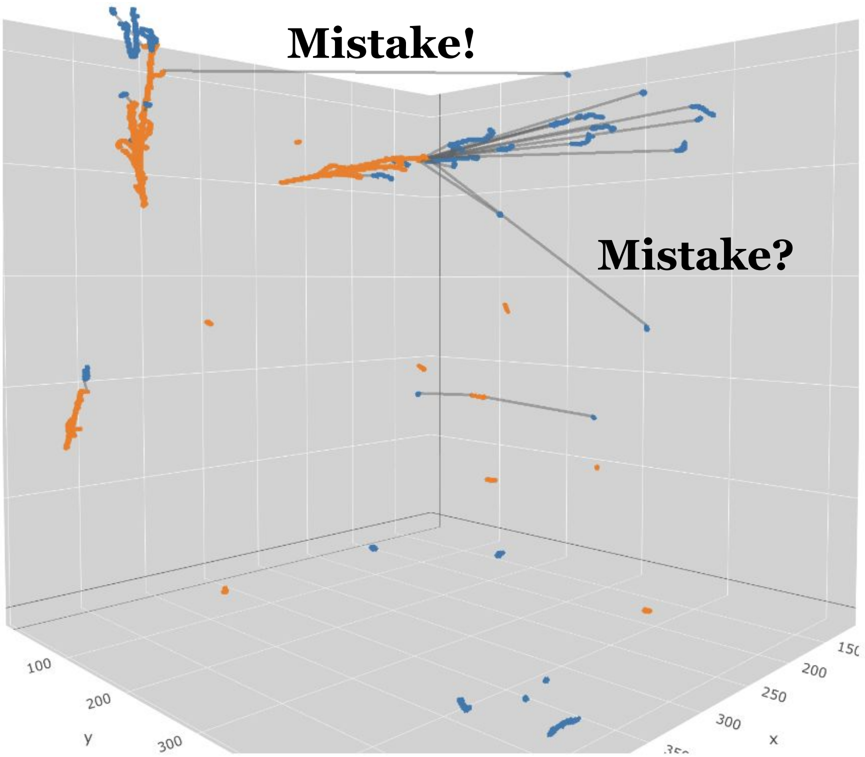
SLAC

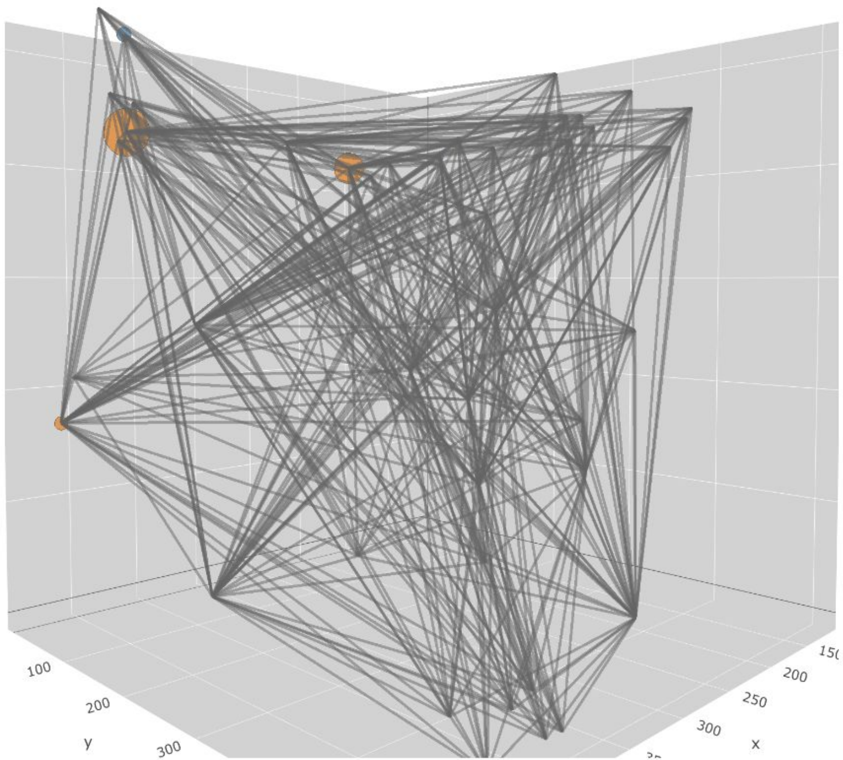
Shower clustering using Bipartite Graph





Shower clustering using Bipartite Graph

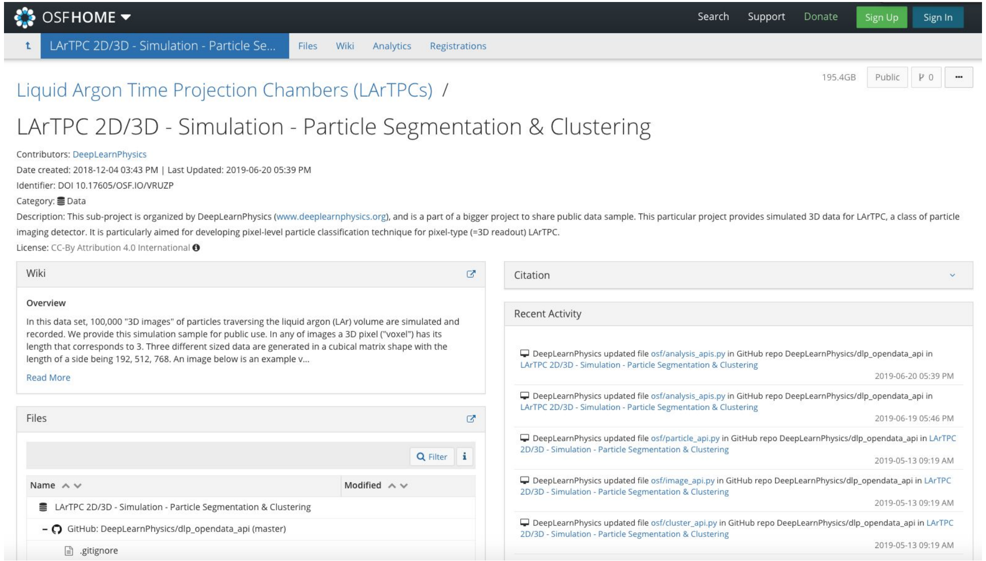


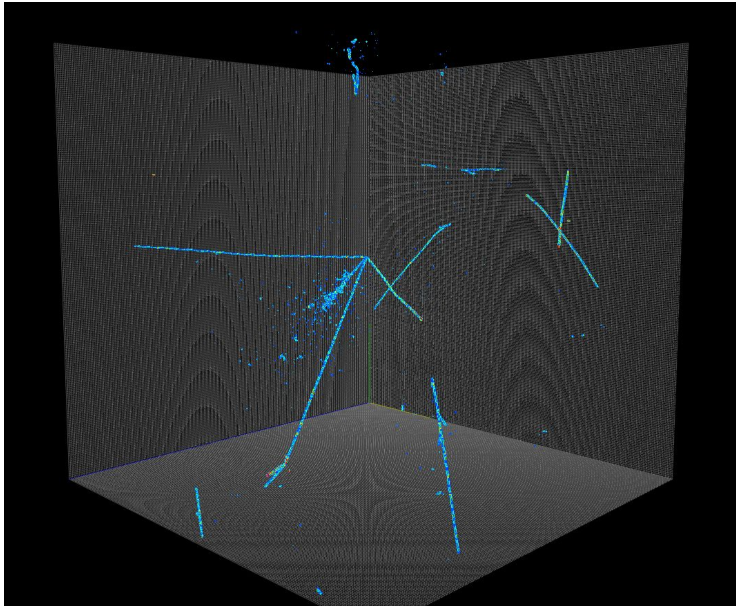


… yep, should wrap-up …
REALLY

Inter-experimental collaborative work

- Open simulation sample
  - **Open real data?** Soon! (3D proto-type R&D @ SLAC)



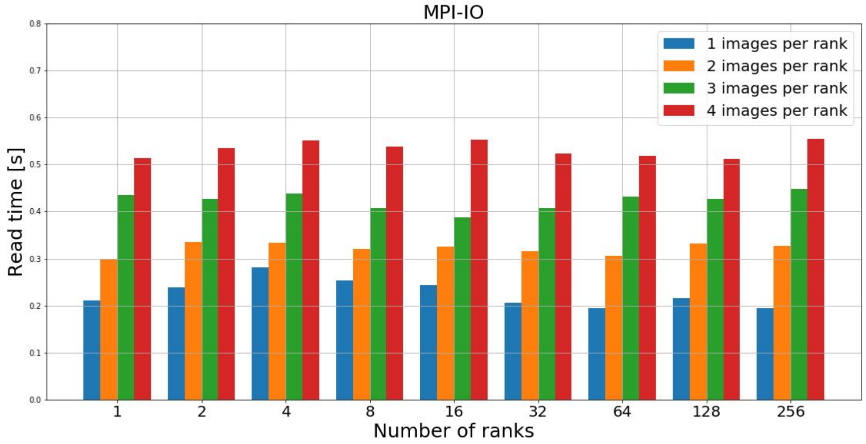


ML-based Neutrino Data Reconstruction Chain
Open Source Development

Inter-experimental collaborative work

- Open simulation sample
  - **Open real data?** Soon! (3D proto-type R&D @ SLAC)
- **Open software development**
  - Fast, distributed IO, optimized for sparse data


Work credit:
Corey Adams (ANL)
Marco del Tutto (FNAL)



- HDF5 format (this doesn't mean anything though)
- Custom format for sparse data for fast IO
- Custom API for data distribution using MPI
  - With horovod, tested hundreds of GPU over servers with infini-band interconnect

Custom development among hobby-coders from SLAC/ANL/FNAL, lead by Corey Adams @ ANL

Open Source Development

Inter-experimental collaborative work

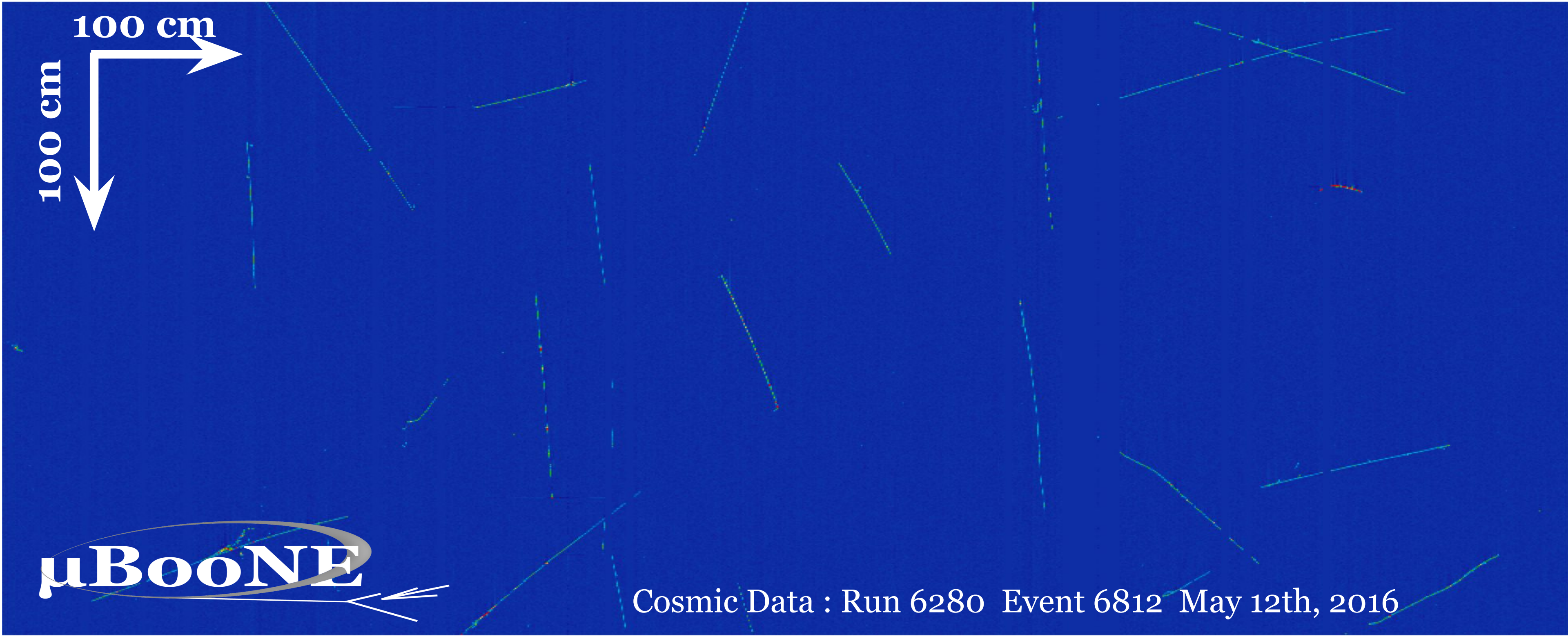
- Open simulation sample
  - **Open real data?** Soon! (3D proto-type R&D @ SLAC)
- **Open software development**
  - Fast, distributed IO, optimized for sparse data
  - GitHub repositories **+** Software container (Singularity/Docker)
  - Example paper w/ data+repo+container ref.
    - Readers reproduced our work!
    - All algorithms shown today in repo.

Closing remarks

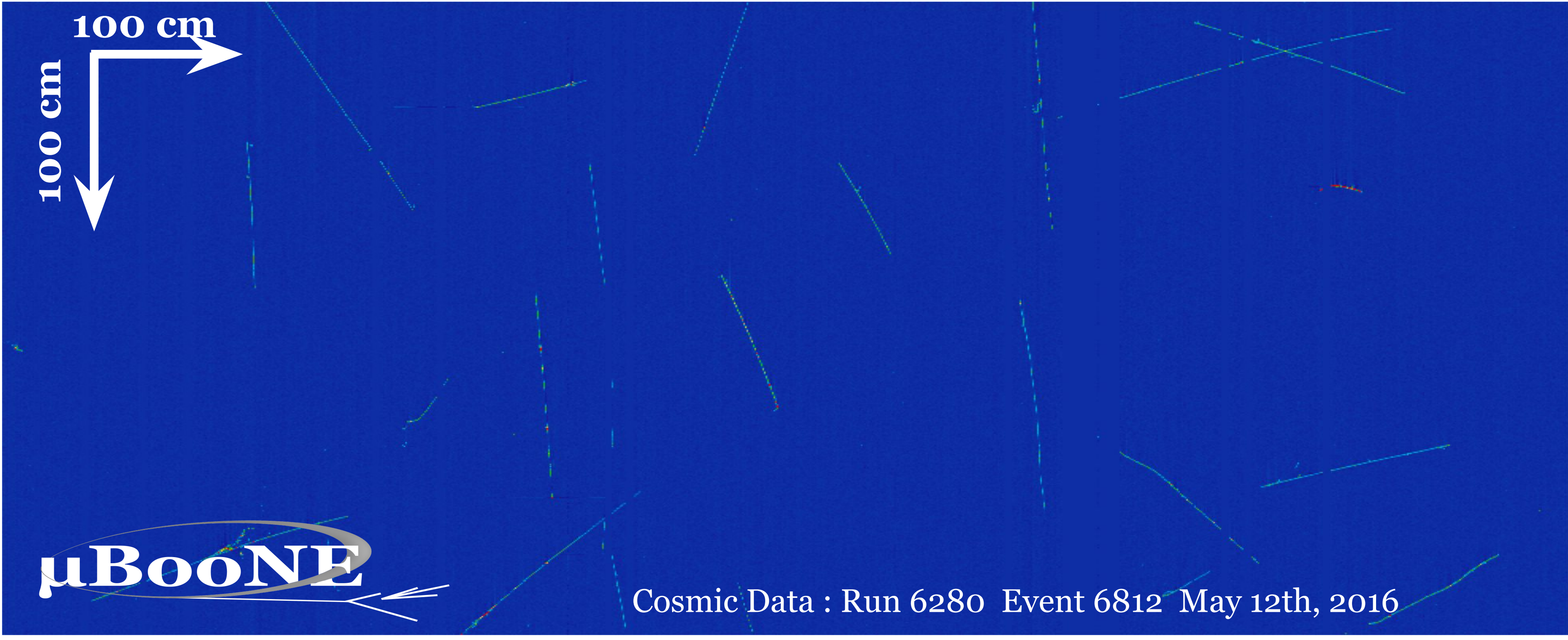
- **Neutrino detector trend: hi-res. particle imaging**
- **Analysis trend: computer vision algorithms**
  - Benefit the hi-resolution image = lots of heuristics (in non-ML)
  - BTW, another cool sparse CNN implementation this summer!
- **ML-based data reconstruction approach**
  - especially for "busy" detectors … my research :)
- **Other active areas**:
  - 3D point reconstruction, data vs. sim. domain discrepancy adaptation, etc. … would love to have a chat!

Questions?

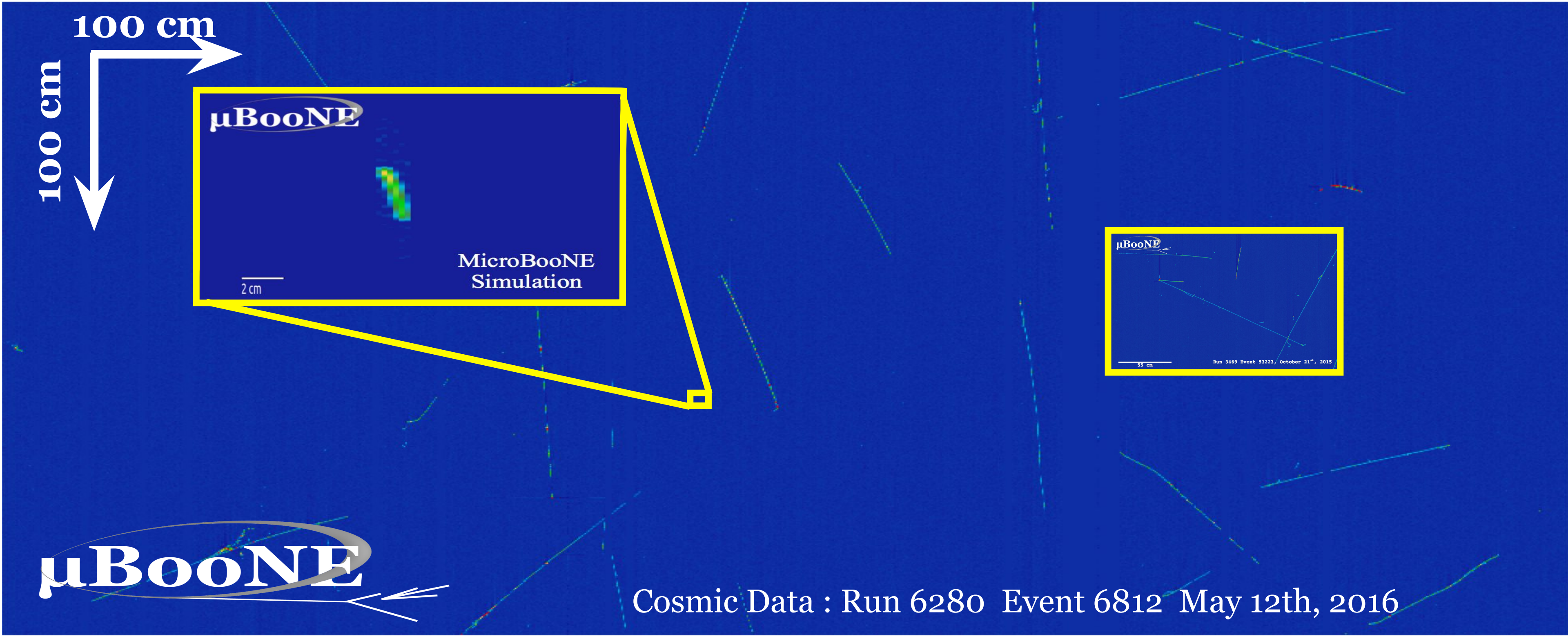
Do you see neutrino interaction here?



Nope :) In this detector, only ~1/700 beam neutrino interacts



... and 1/600 have many variations in hi-resolution imaging...



Submanifold Sparse Convolutions

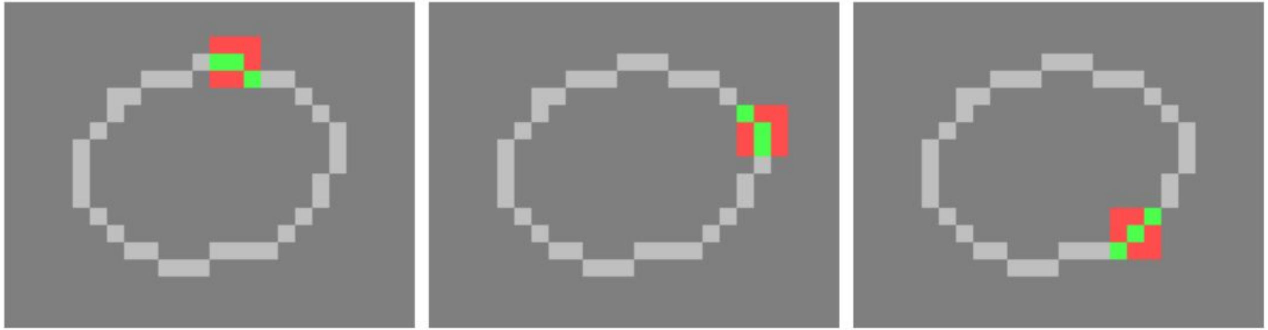
1. **Resources waste** of dense convolutions on sparse data
2. **Dilation problem**

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

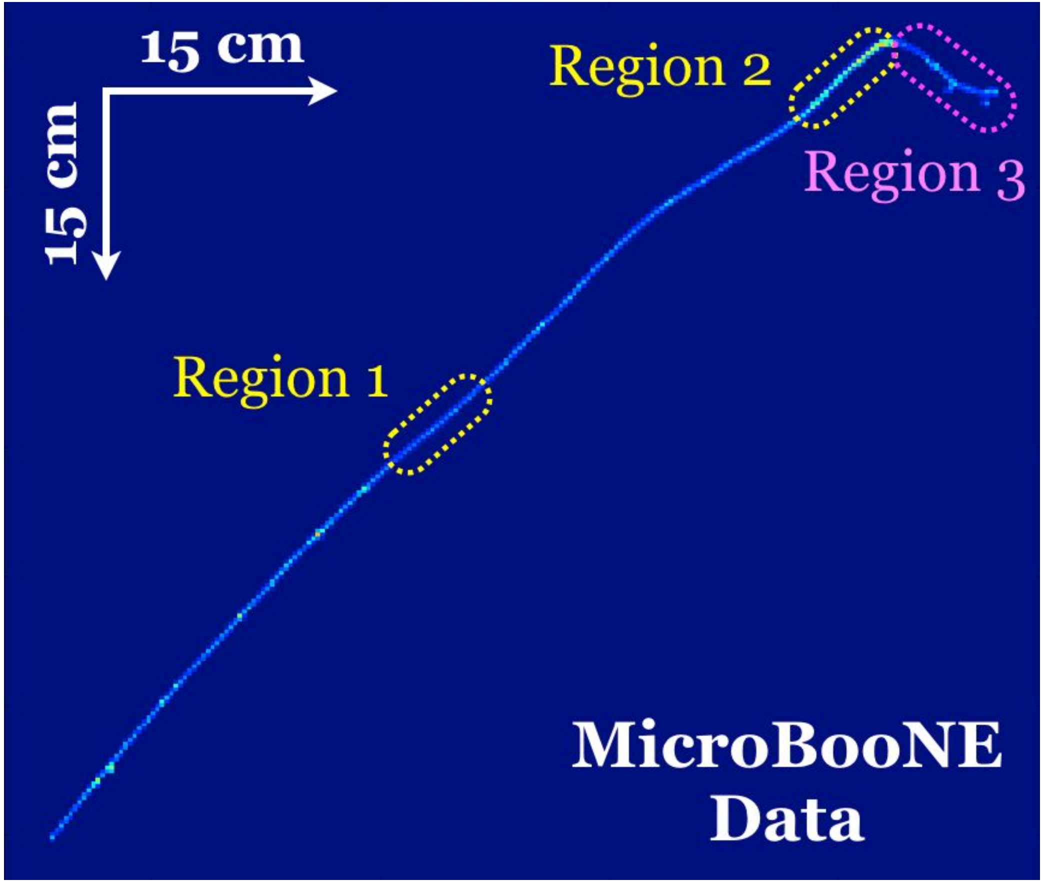
3D Semantic Segmentation with Submanifold Sparse Convolutional Networks (arxiv: 1711.10275)

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

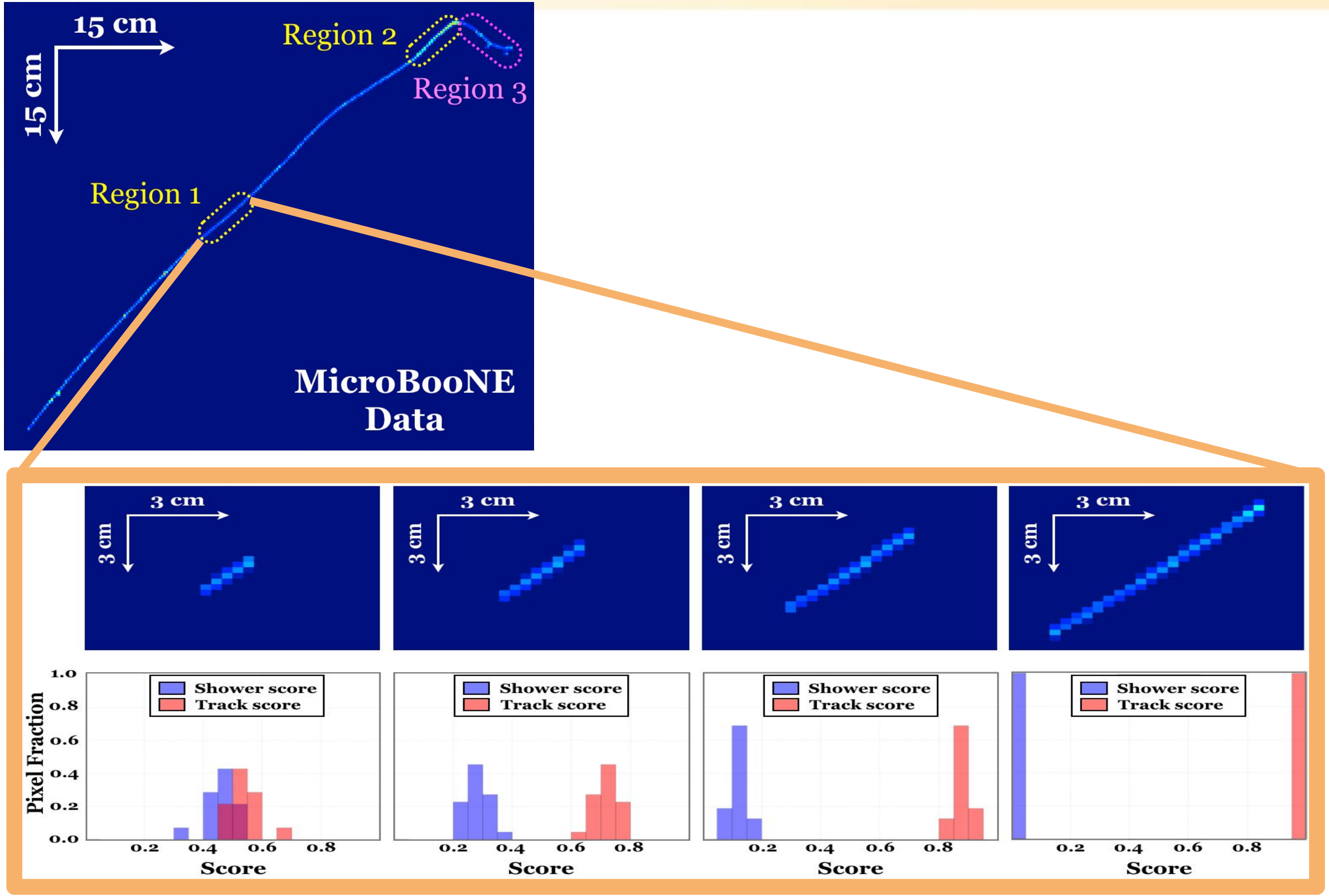


Fun Playing with Semantic Segmentation



Machine Learning & Computer Vision in Neutrino Physics

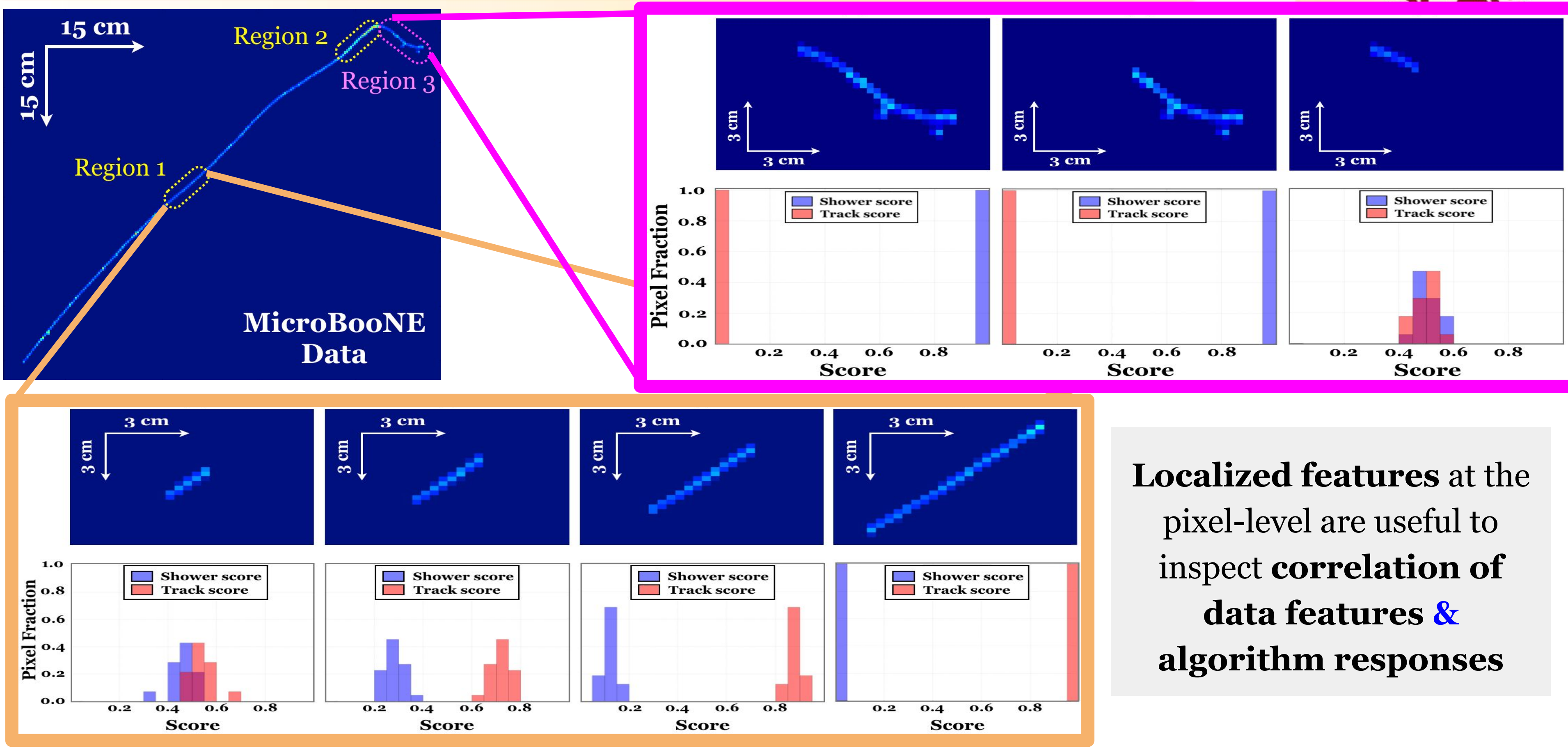
Fun Playing with Semantic Segmentation



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**