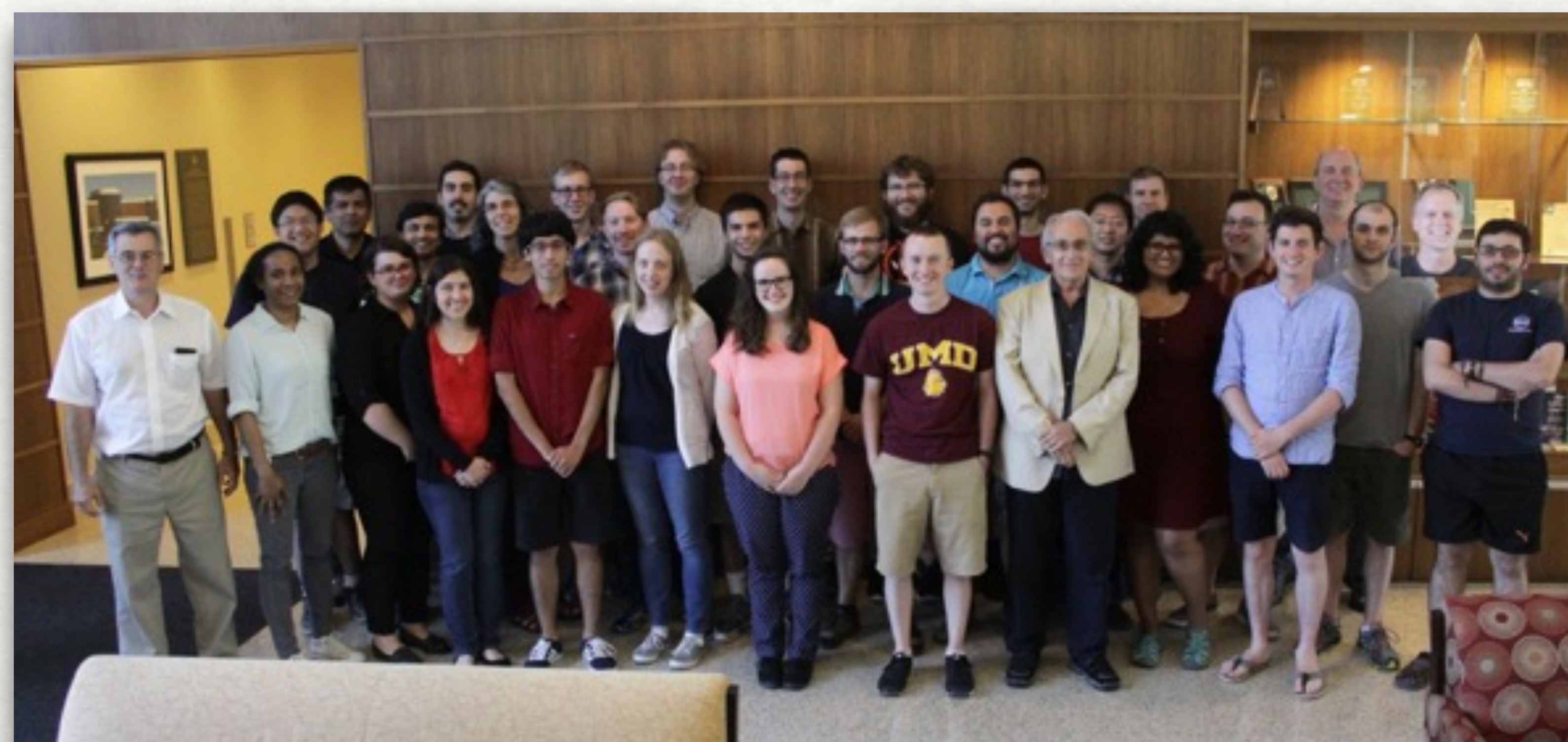
JONATHAN MILLER
UNIVERSIDAD TECNICA FEDERICO SANTA MARIA
FOR THE MINERVA COLLABORATION

EXPLORATION OF DEEP CONVOLUTIONAL AND DOMAIN ADVERSARIAL NEURAL NETWORKS IN MINERVA.

ACKNOWLEDGEMENTS

THE MINERVA COLLABORATION

- The MINERvA Collaboration is a productive collaboration of ~60 physicists from ~20 institutions in ~10 countries.
- This is the work of MINERvA Machine Learning working group and as such primarily the work of it's fearless leader Gabriel Perdue and it's students and postdocs: Marianette Wospakrik, Anushree Ghosh and Sohini Upadhyay.



CHALLENGE FOR ANALYSIS IN PARTICLE PHYSICS THIS CENTURY

UNIQUE CHALLENGES

- Particle physicists (MINERvA, IceCube, etc) produce an enormous amount of data:
  - Detectors with many channels create a high resolution image of event
  - Astrophysics and particle physics are often in the ``intensity frontier" with an enormous data rate
- Previous century: Scanners (photographic plates), counting and simple `bottom up' algorithmic procedures
- This century: Machine Learning and Pattern Recognition

Brookhaven National Lab

CHALLENGE FOR ANALYSIS IN PARTICLE PHYSICS THIS CENTURY

PERSONAL QUEST SINCE 2012

- The amount of data, due to the size of the detectors and the number of relevant events, poses unique challenges:
  - Difficult to `find' the most useful variables (or features)
  - Simulation (or `labeled data') is required for analysis but may have `artifacts' which do not exist in data (which is `unlabeled').
- Machine Learning Algorithms are complicated
  - Support Vector Machine or Boosted Decision Tree or Neural Network or k-Nearest Neighbors and then training speed, parameters, kernel, kernel properties, layers, etc?

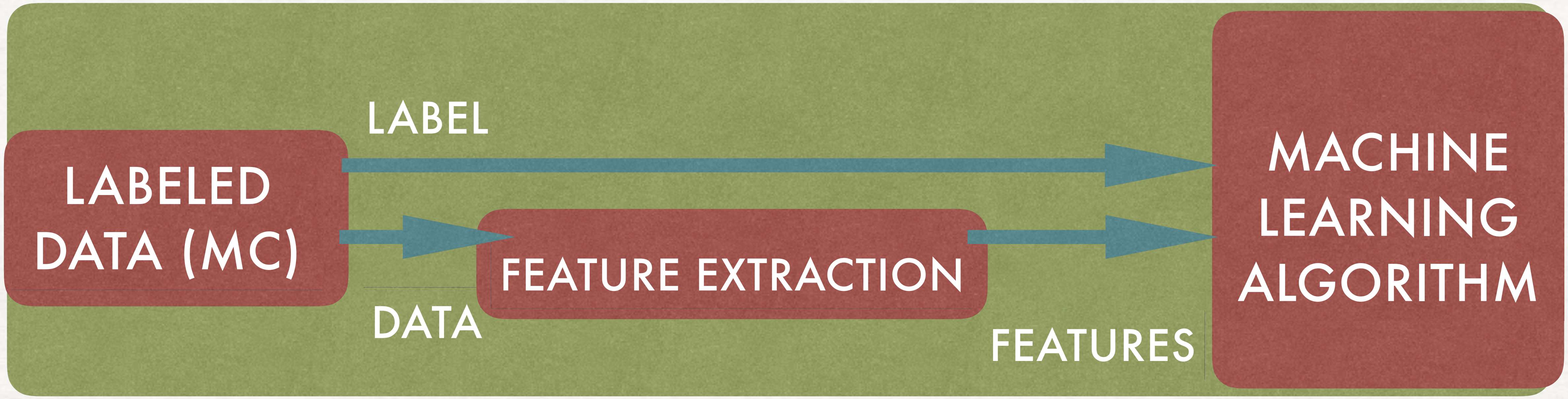
CHALLENGE FOR ANALYSIS IN PARTICLE PHYSICS THIS CENTURY

NEW DIRECTIONS FROM COMPUTER VISION

- Challenge due to volume of data:
    - Follow lead of computer vision and pattern recognition and use Convolutional Neural Networks to extract geometric features.
        - This was enabled by the advent of GPUs and algorithmic advances (dropout, initialization, etc). - Revolutionary
    - Development of domain-adversarial training as solution to having lots of unlabeled data but little labeled data (arXiv:1505.07818).
- Complexity of MLA: use Neural Networks
    - See talks next year about optimizing topology/parameters (ALCC HEP 109 at ORNL).

MACHINE LEARNING ALGORITHMS (MLA)

CARTOON

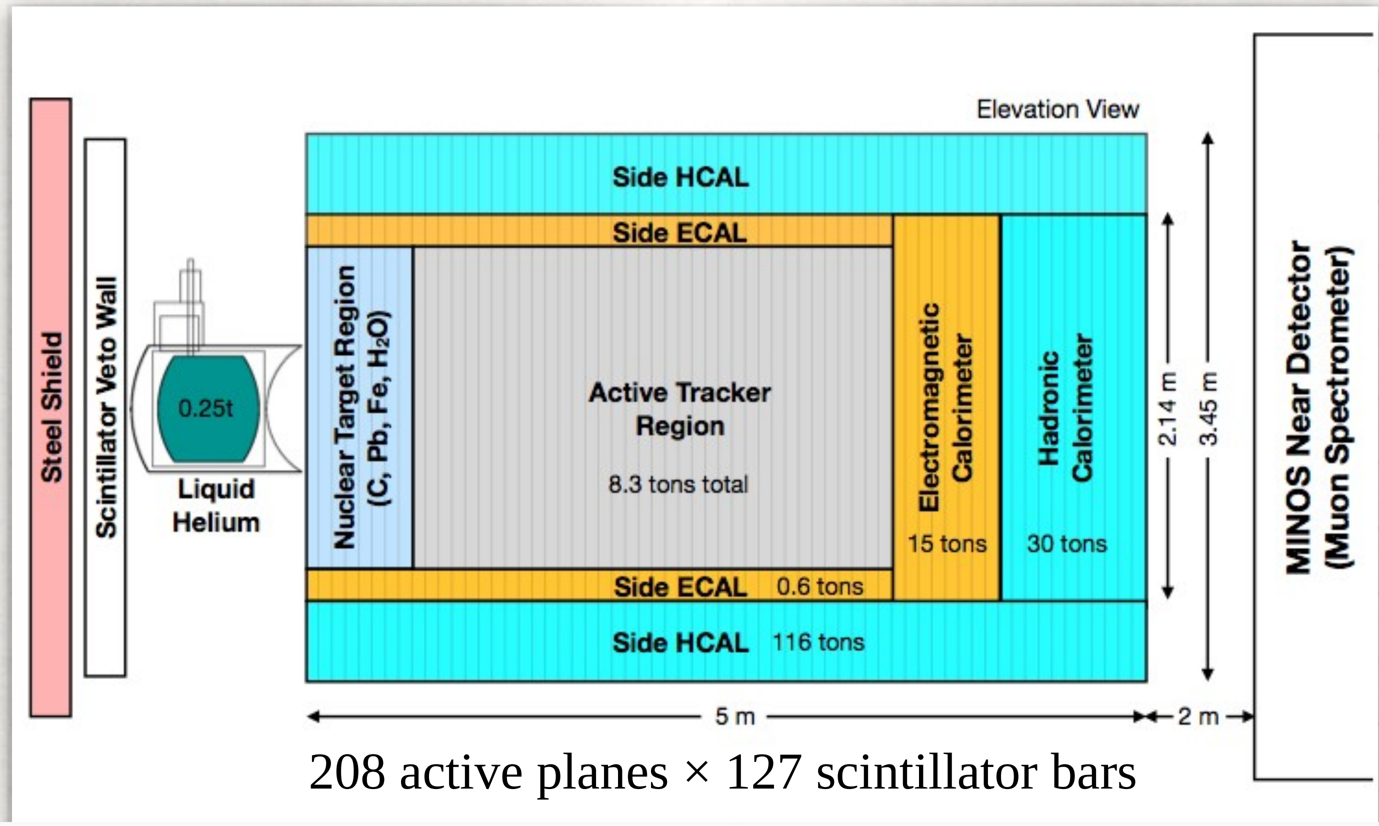


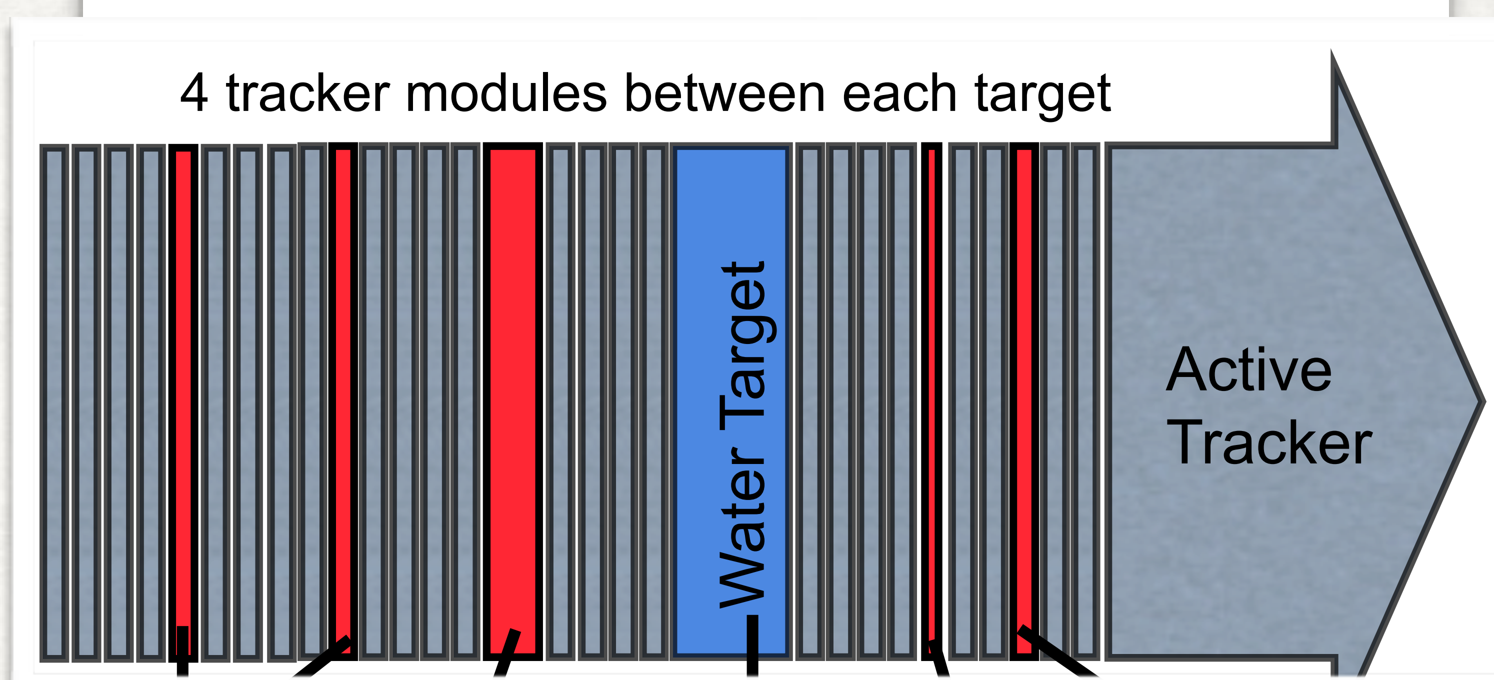
- Feature extraction realized by procedural algorithms. (Human Intelligence)
- MLA can provide new variables which can then be fed into later MLA.
- Developing and selecting variables and features to feed into a well behaved and high impact MLA is one of the greatest challenges in an analysis.

MINERVA EXPERIMENT AT FERMILAB

HIGH RESOLUTION IMAGE

- 120 modules for tracking and colorimetry (32k readout channels)
- The MINOS near detector serves as a muon spectrometer.
- Made up of planes of strips in 3 orientations: X, U, and V.
- Includes Helium target, water target and 5 passive nuclear targets made up of Carbon, Iron and Lead.

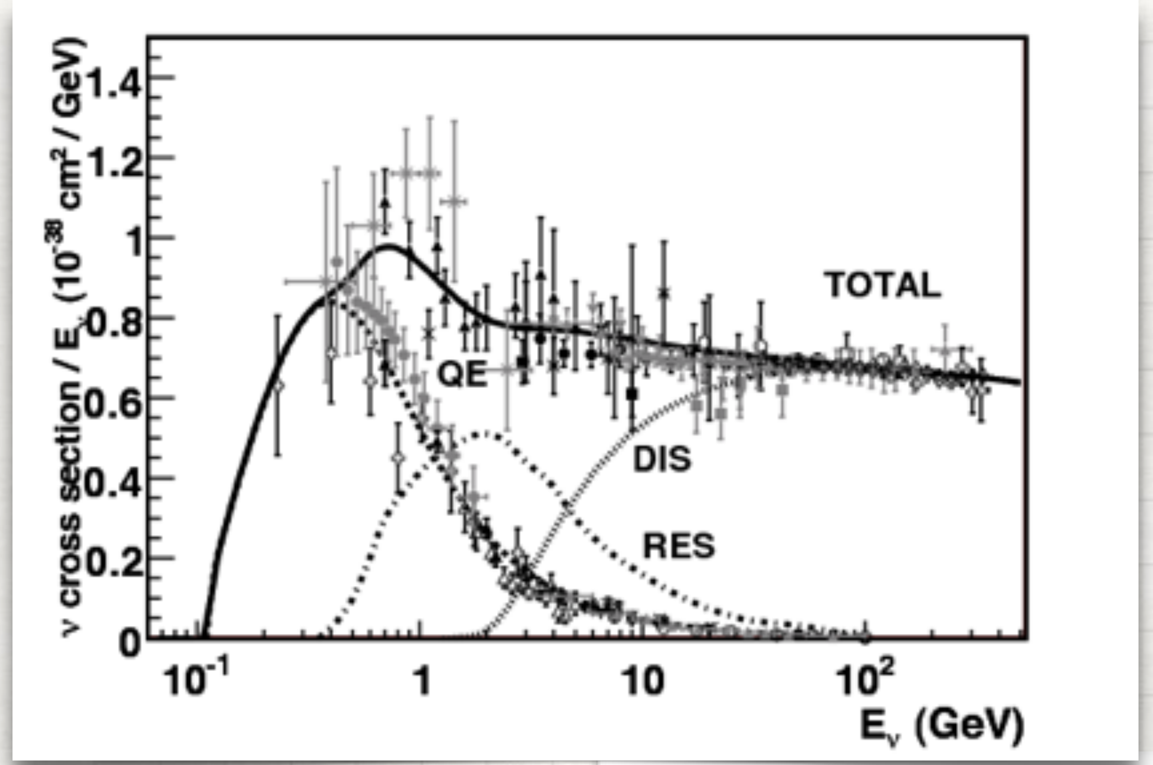


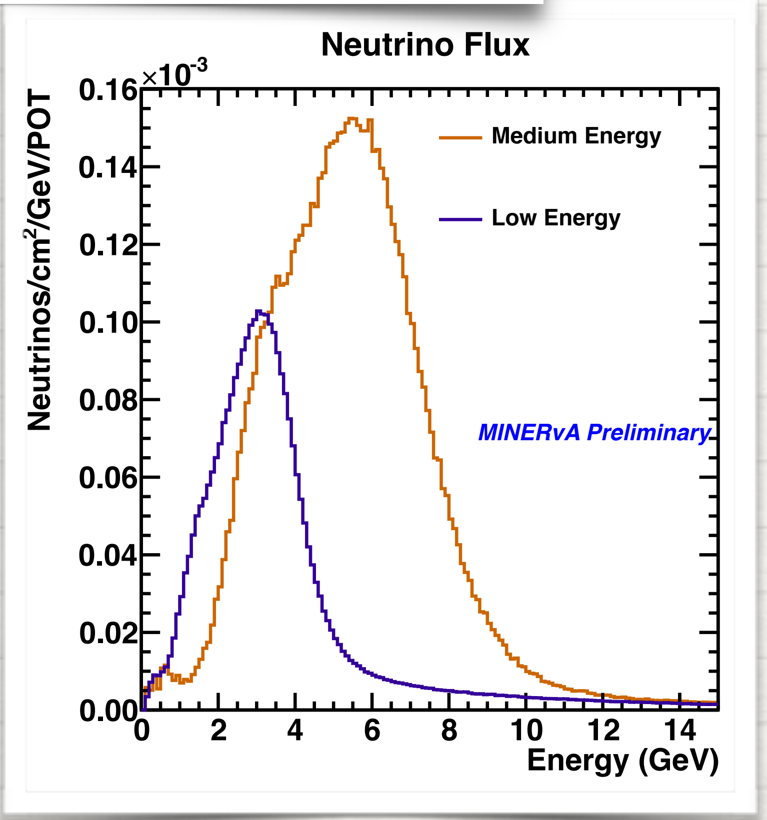


MINERVA EXPERIMENT AT FERMILAB

LOTS OF DATA AND COMPLICATED IMAGE

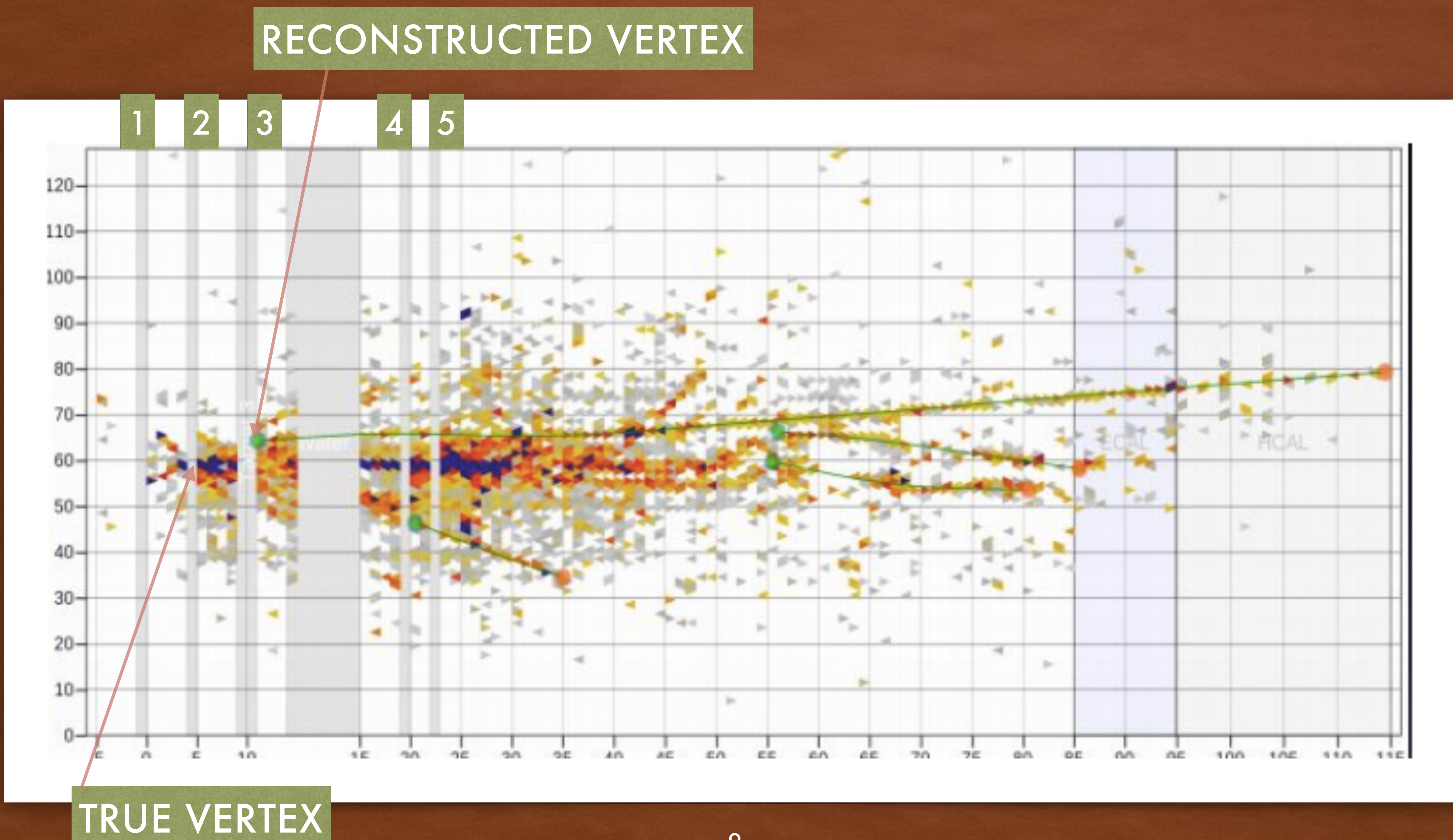
- We have taken 12E20 Protons-On-Target in the Medium Energy (ME) neutrino beam (6E6 in one playlist).
- The higher statistics and energy means improved neutrino nuclear measurements.
- The majority of the flux is now in the DIS region. Deep Inelastic Scattering is a more challenging reconstruction.





IN DIS EVENTS LARGE AND COMPLICATED HADRONIC SHOWERS MAY MASK THE PRIMARY VERTEX FROM TRACK BASED ALGORITHM (WALK BACK PRIMARY TRACK AND LOOK FOR SECONDARIES)

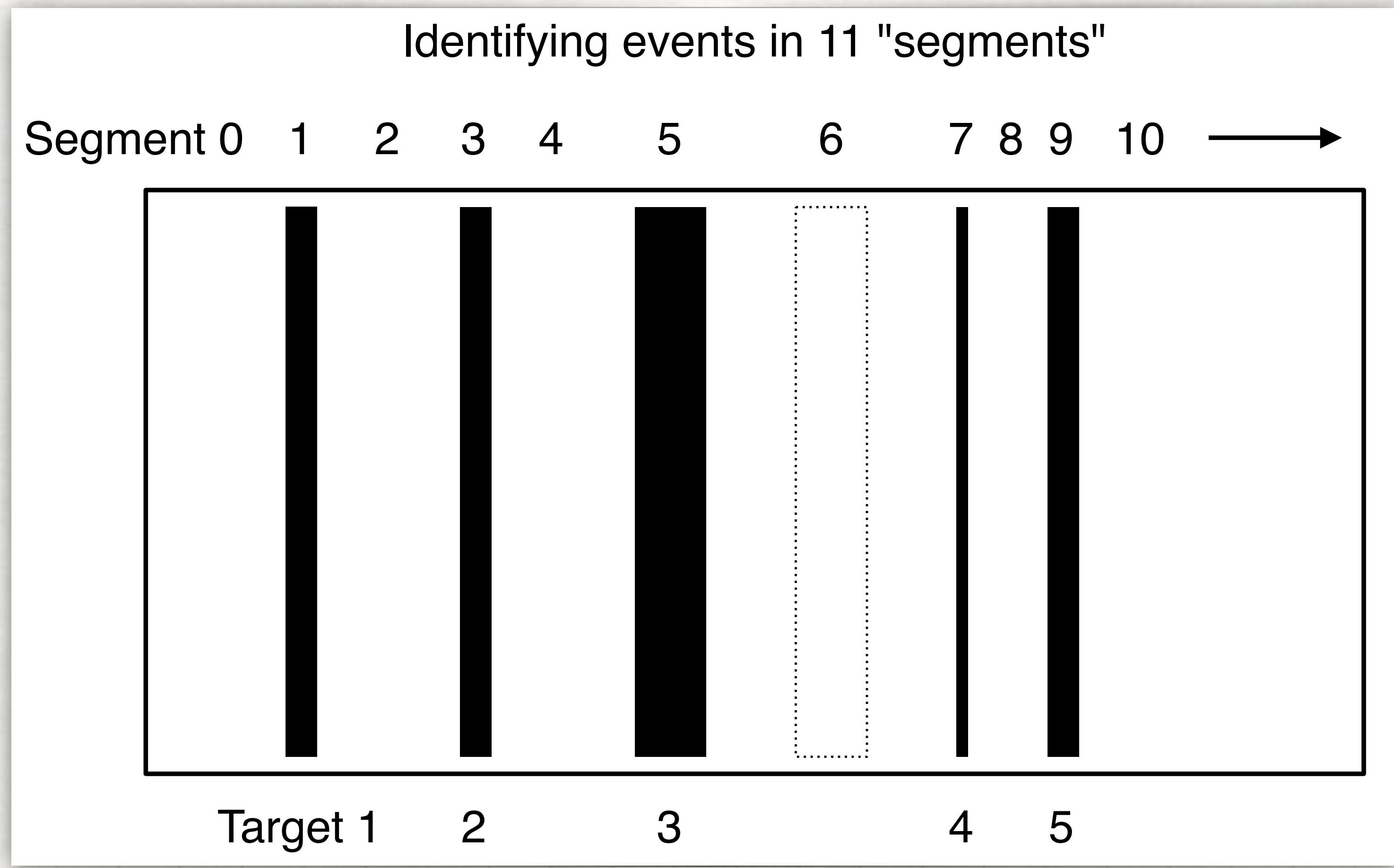
MINERVA VERTEX FINDING

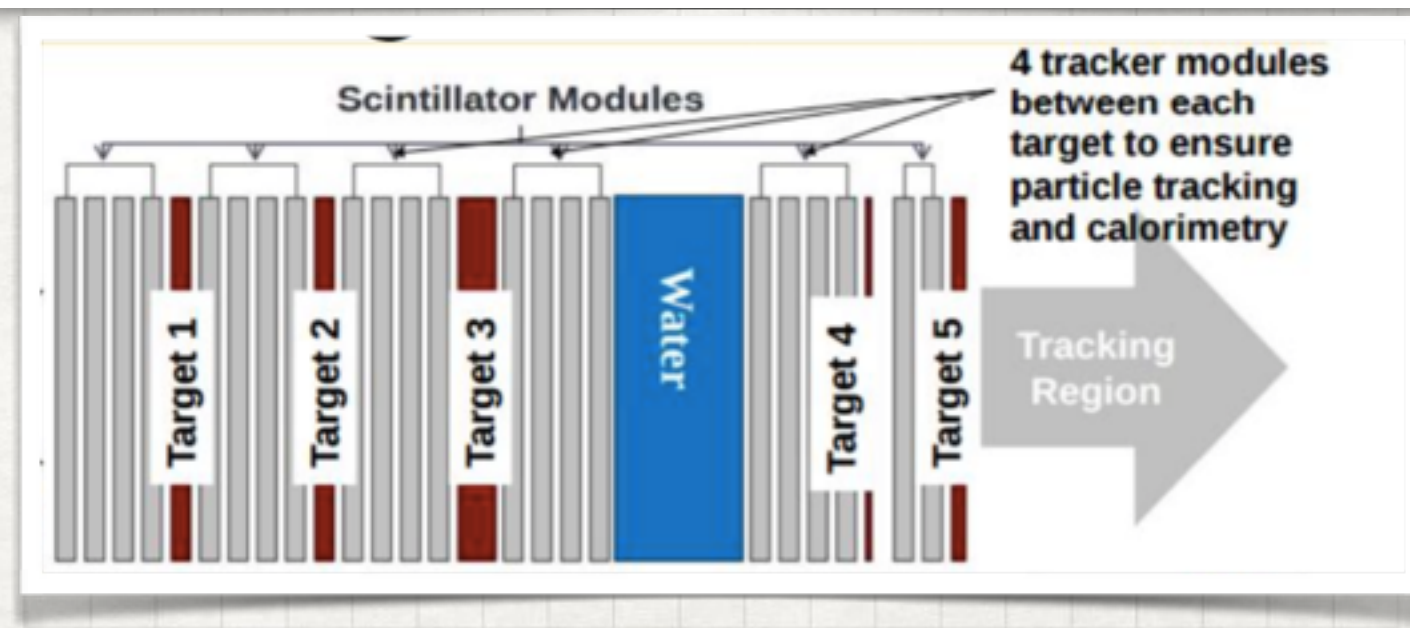


MINERVA VERTEX FINDING

DEEP NEURAL NETWORK (DNN)

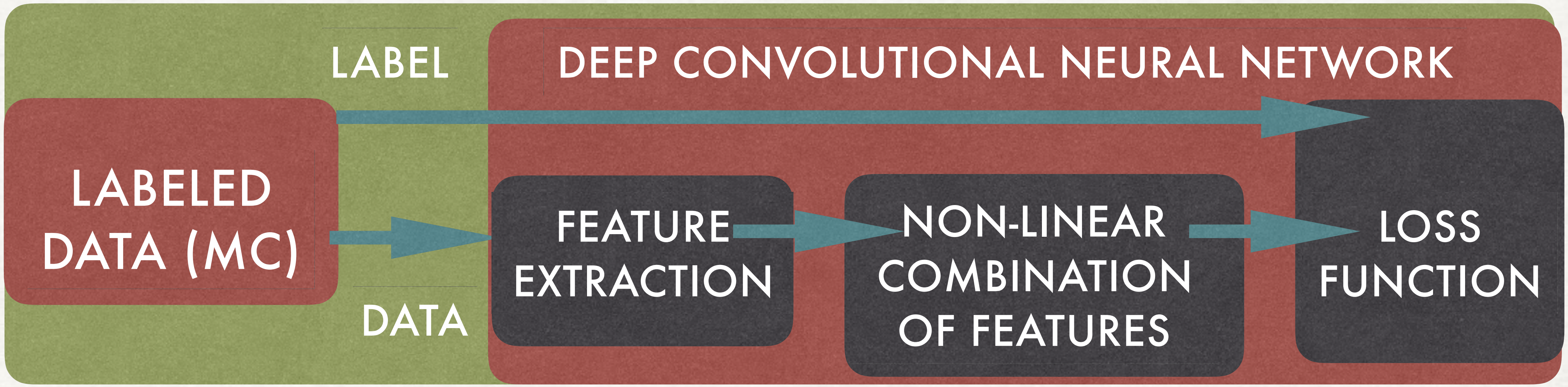
- DNN provides prediction of the segment (or plane number) an interaction is in.
  - We use non-square kernels and pool along X,U,V to collapse into semantic space in X,U,V but leave z unchanged.
  - Plane number is done the same but class is based on plane number and not segment.
- Between targets only 2 (1 in segment 8) pixels in U and V.
- Only the first planes of downstream is included in segment 10.





MACHINE LEARNING

DEEP CONVOLUTIONAL NEURAL NETWORKS



- Feature extraction is realized within the MLA. This extraction may be convolved with the nonlinear construction of more complicated features and optimization.

- Convolutional Neural Network may be used only for feature extraction.

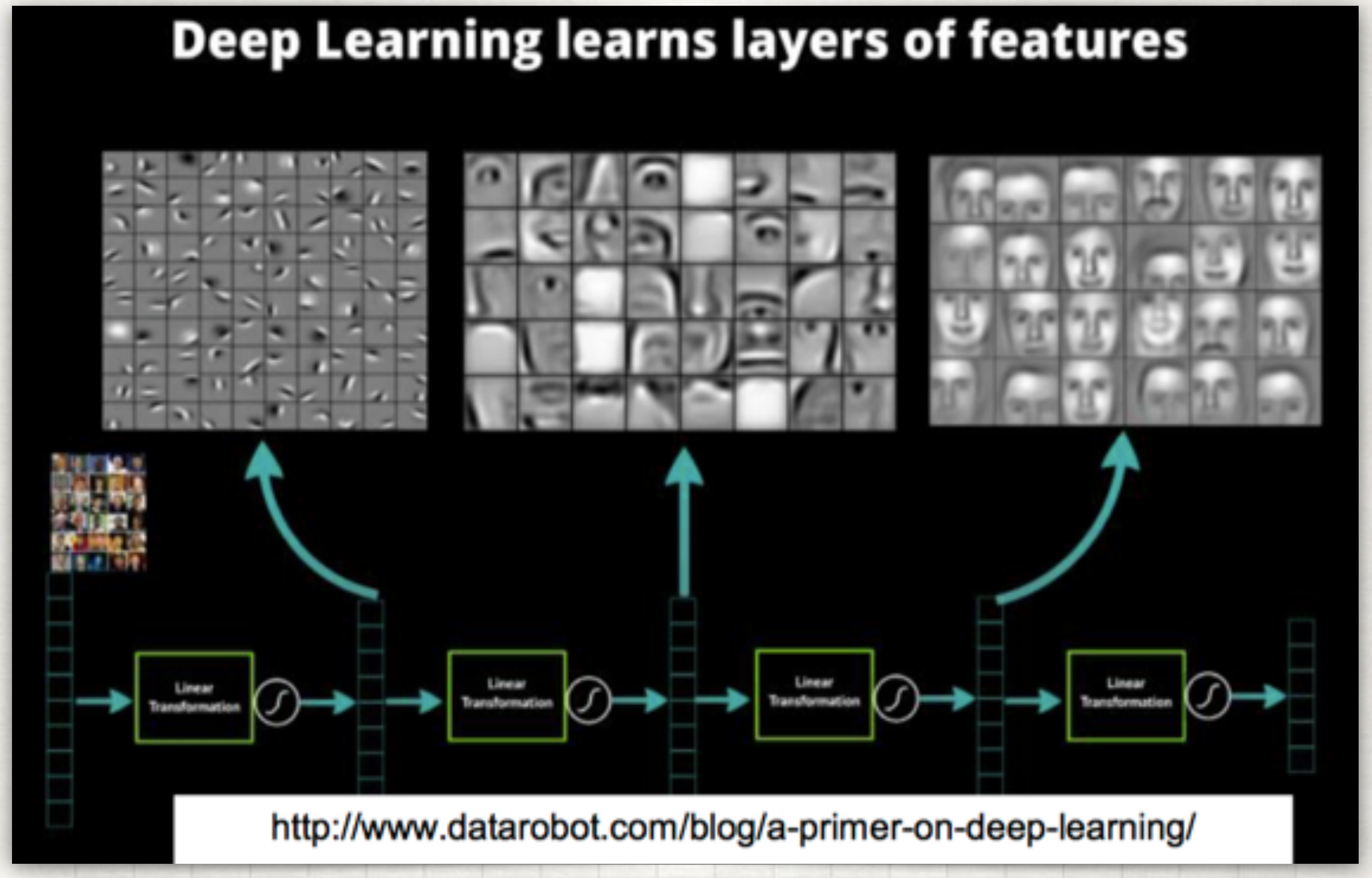
DEEP NEURAL NETWORK (DNN)

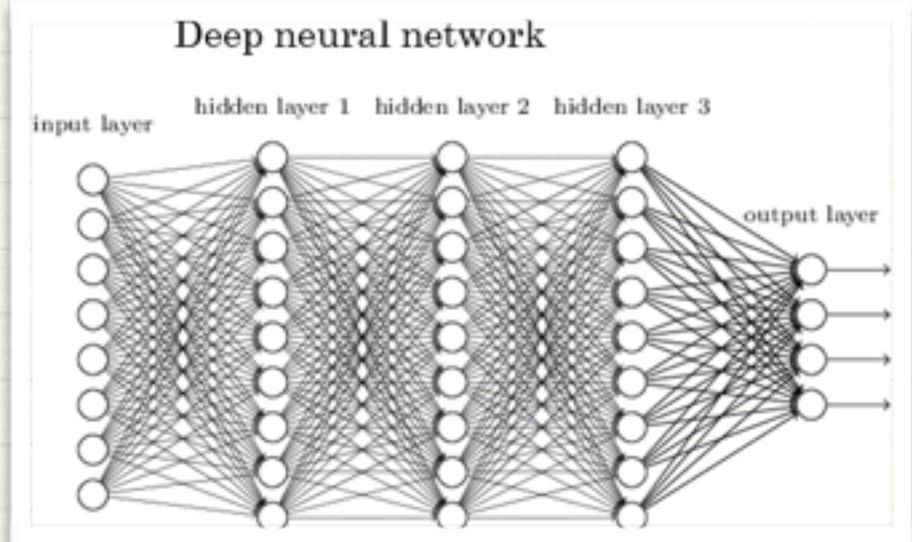
NONLINEAR FEATURE EXTRACTION

- This is the `hierarchal model' where the representations in early layers are combined in the later layers.

- A deep system of nonlinear layers and fully connected layers allow for the production of complicated nonlinear combinations.

- In a deep neural network, the early layers of the network `learn' local features while the later layers `learn' global features.

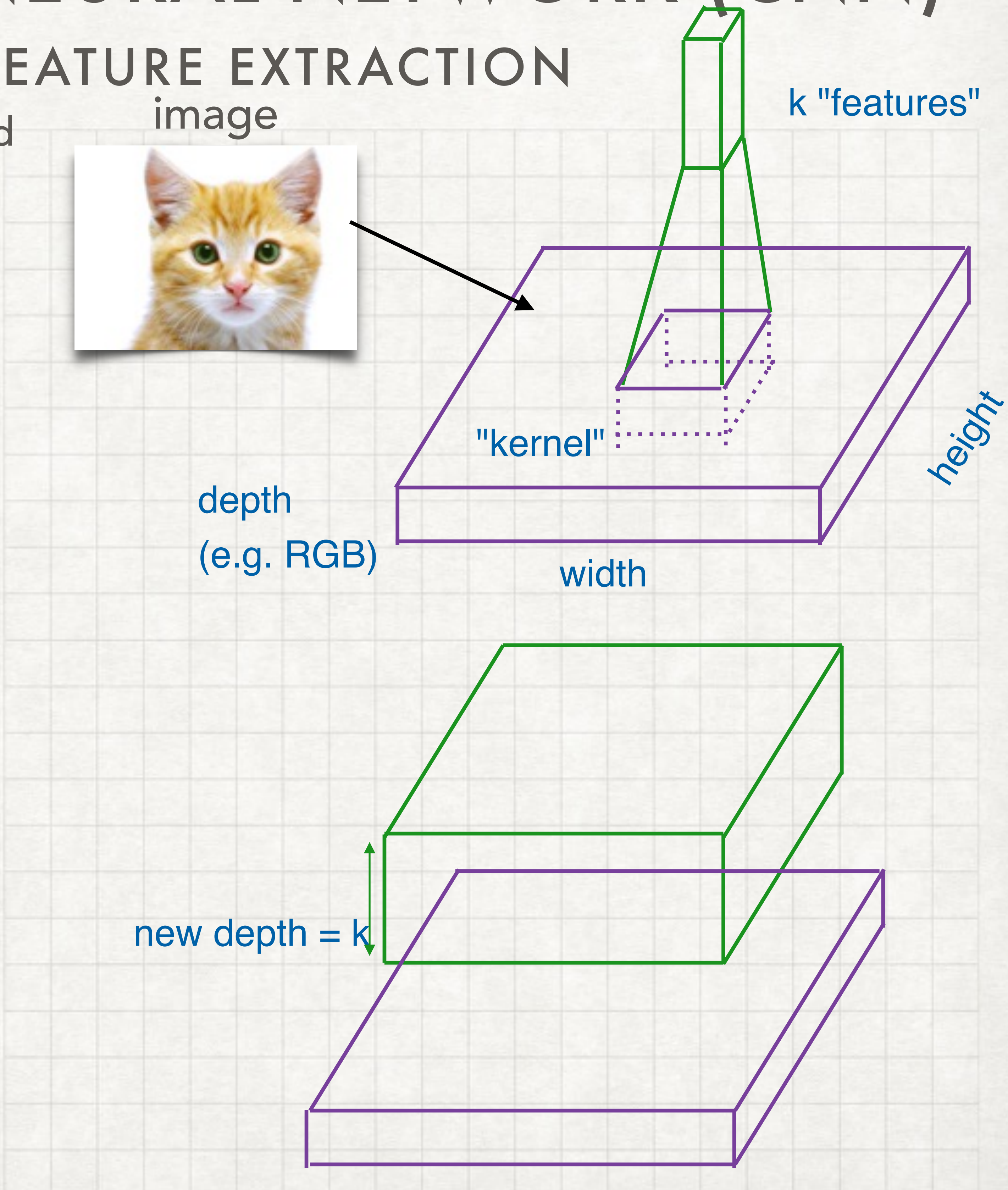


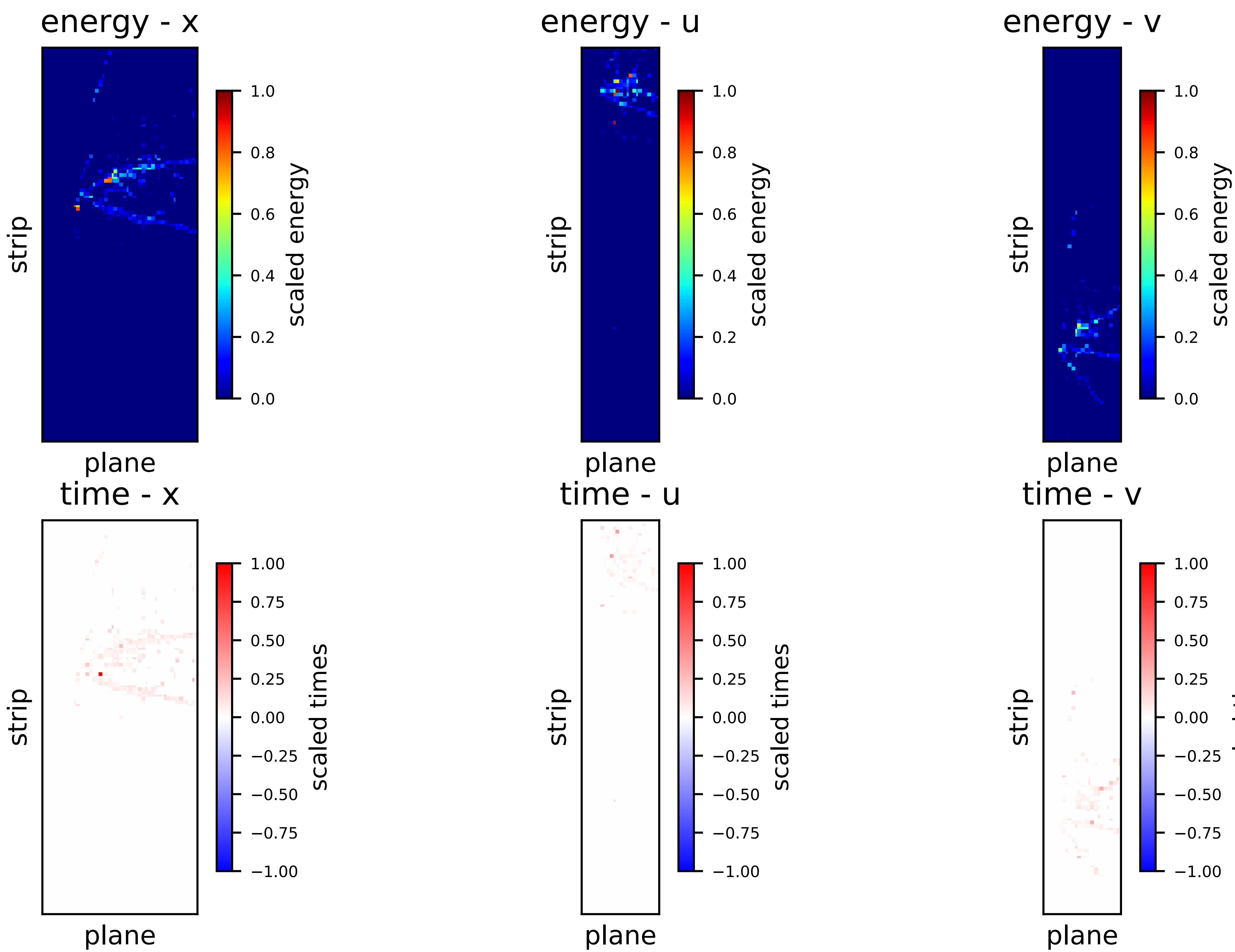


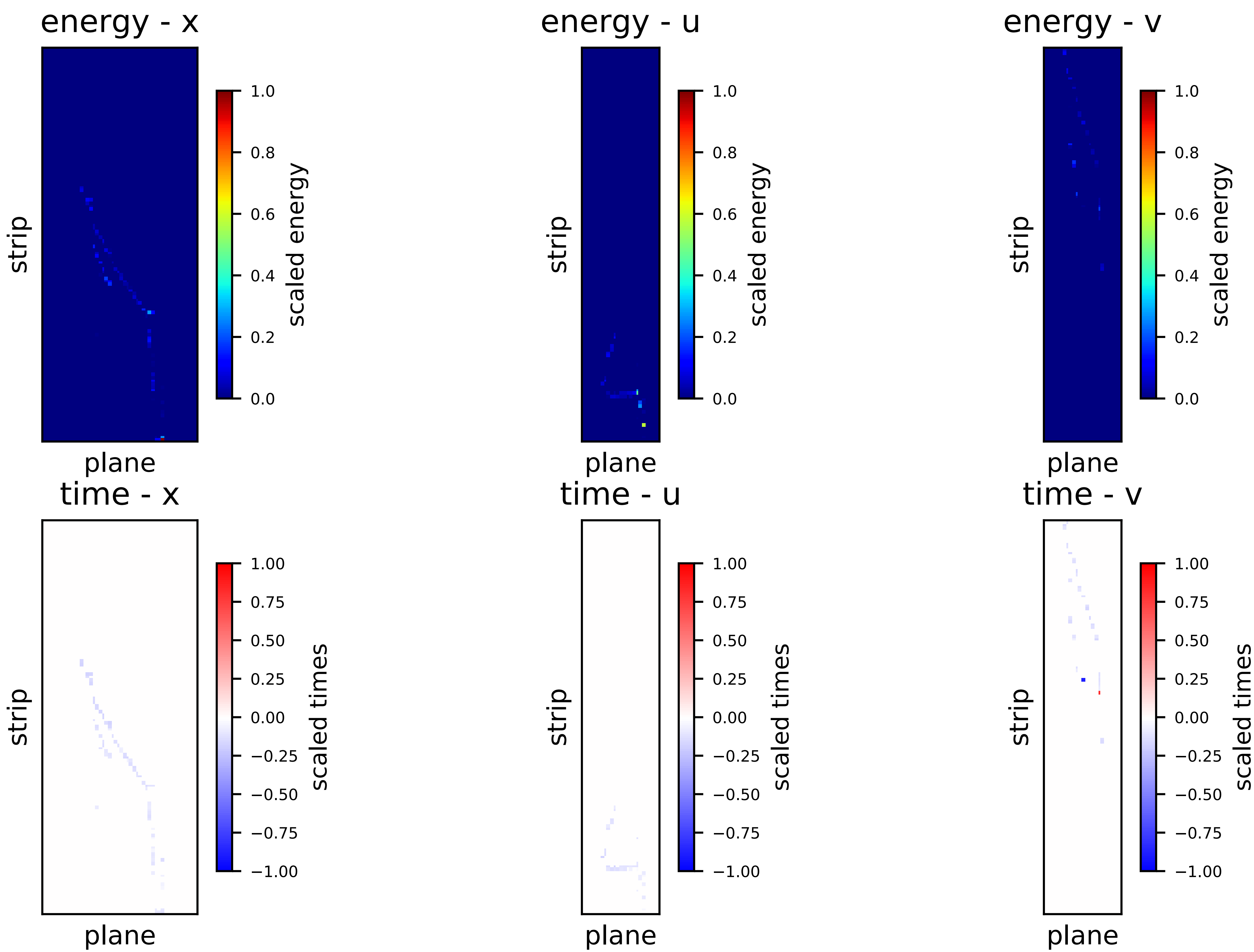
CONVOLUTIONAL NEURAL NETWORK (CNN)

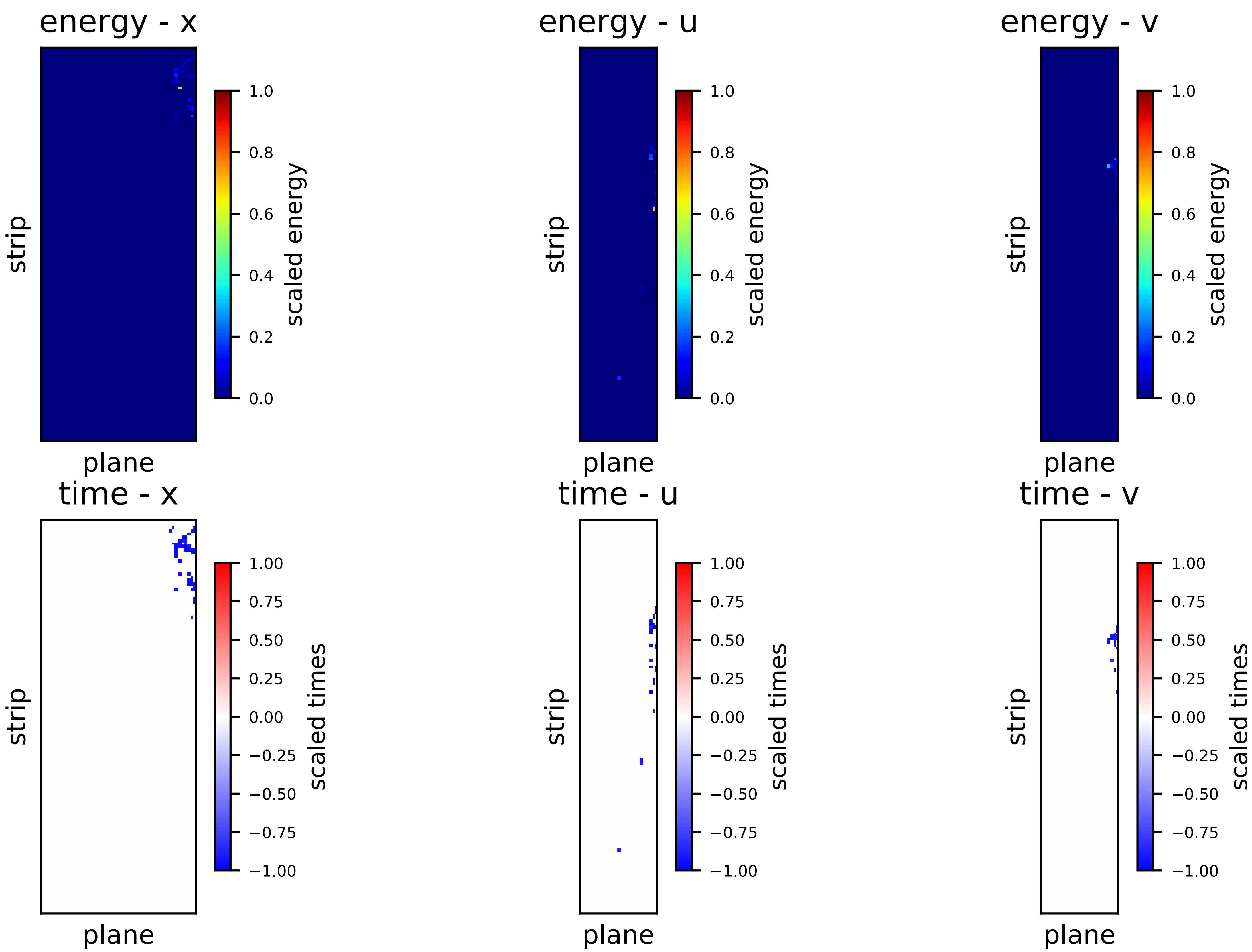
GEOMETRIC FEATURE EXTRACTION

- These types of networks are well suited for feature extraction for things like images with geometric structures.
  - Particle physics events have geometric structures which are procedural algorithms (or scanners) identify.
- Convolutional networks have fewer parameters that are fit due to having only a single parameter across the space (for a given kernel). Parameters describe how the kernel is applied.
- In MINERvA we have time and energy information (obvious use of `depth')
- Final convolutional layer is a `semantic' representation rather than a spatial representation.




energy - x
energy - u
energy - v
scaled energy
1.0
0.8
0.6
0.4
0.2
0.0
strip
plane
time - x
time - u
time - v
scaled times
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00


energy - x
energy - u
energy - v
strip
plane
scaled energy
1.0
0.8
0.6
0.4
0.2
0.0
time - x
time - u
time - v
strip
plane
scaled times
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00


energy - x
energy - u
energy - v
strip
plane
scaled energy
1.0
0.8
0.6
0.4
0.2
0.0
time - x
time - u
time - v
scaled times
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

DEEP CONVOLUTIONAL NEURAL NETWORK FOR VERTEX FINDING

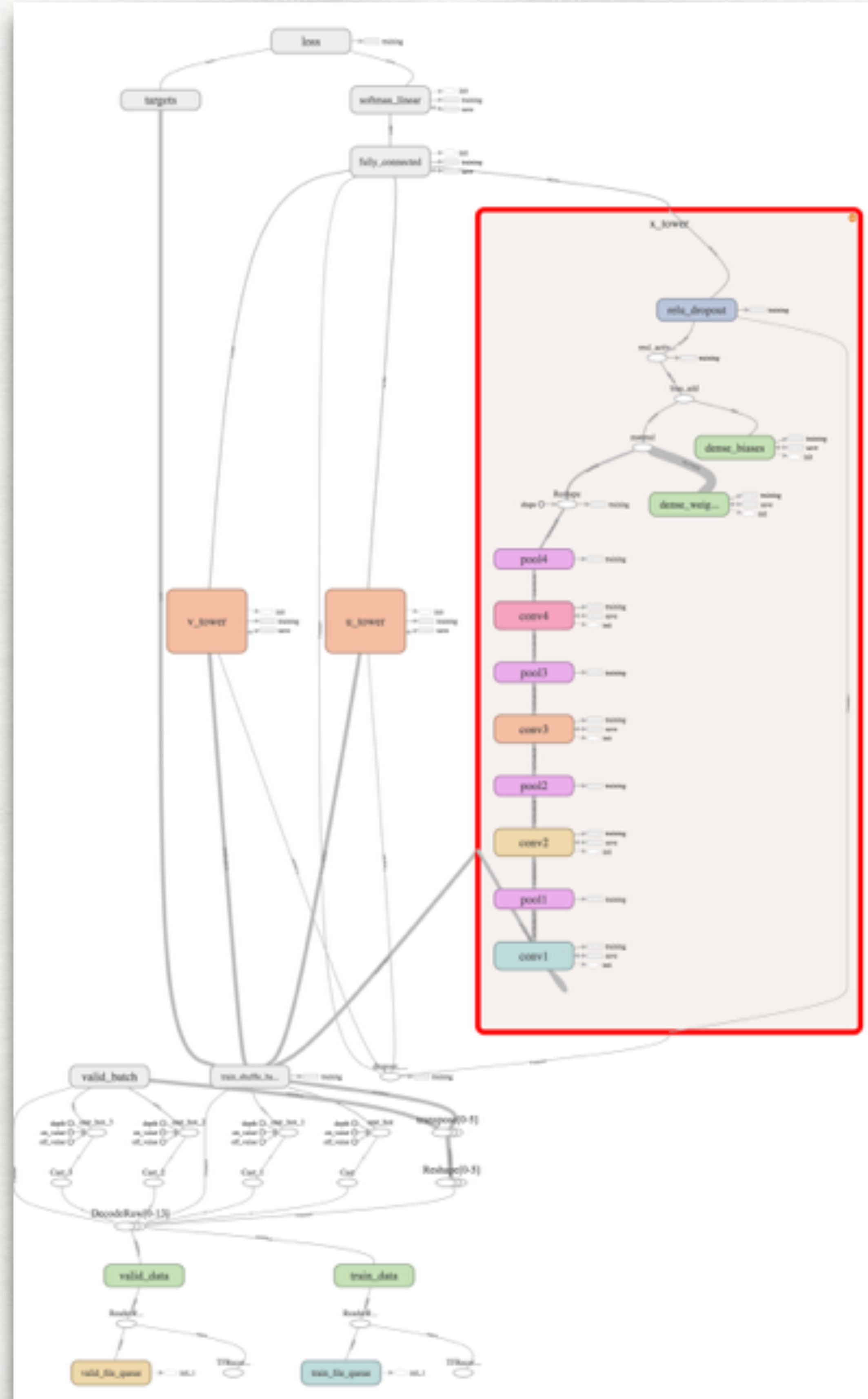
DCNN

- Started from minimalist model and added layers and adjusted filters following intuition.

- We have three separate convolutional towers that look at each of the X, U, and V images.

- These towers feature image maps of different sizes at different layers of depth to reflect the different information density in the different views.

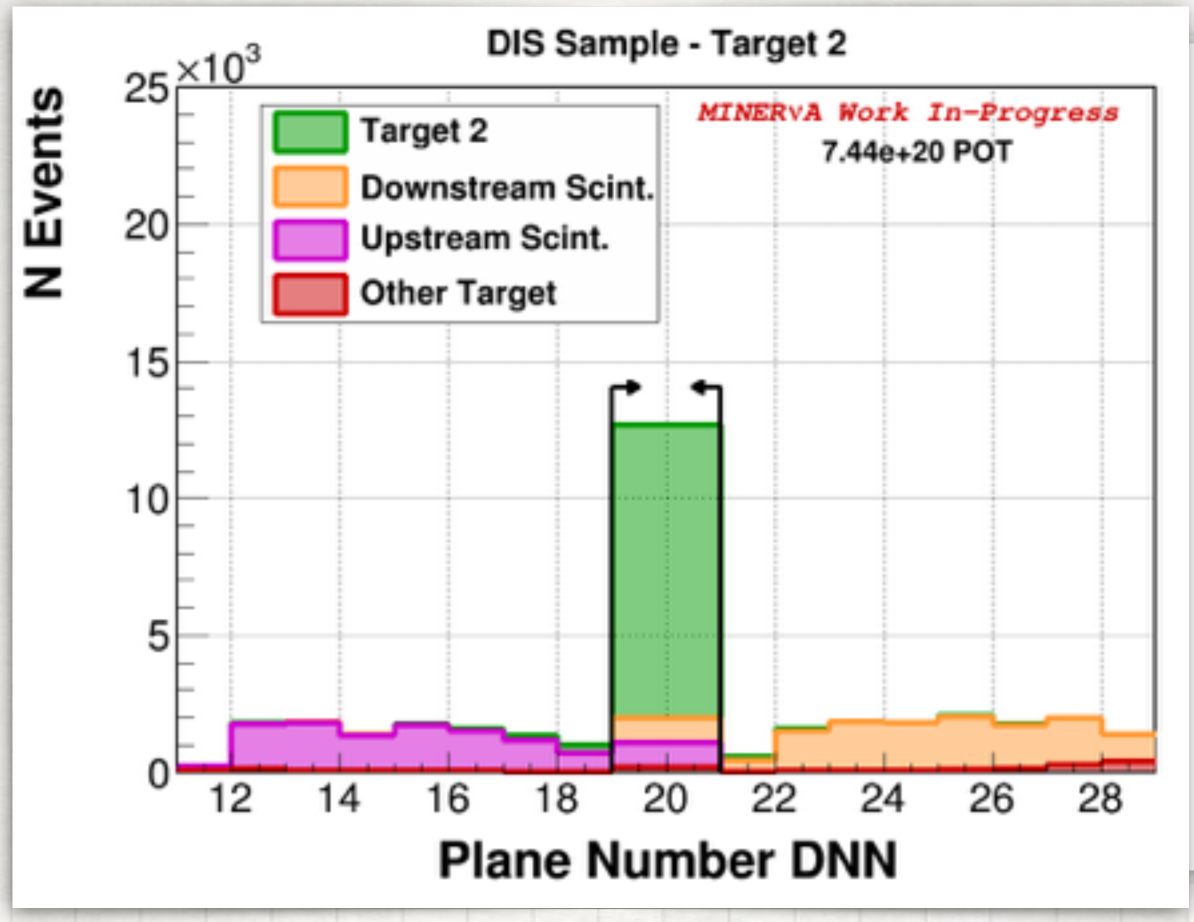
- The output of each convolutional tower is fed to fully connected layer, then concatenated and fed into another fully connected layer before being fed into the loss function.

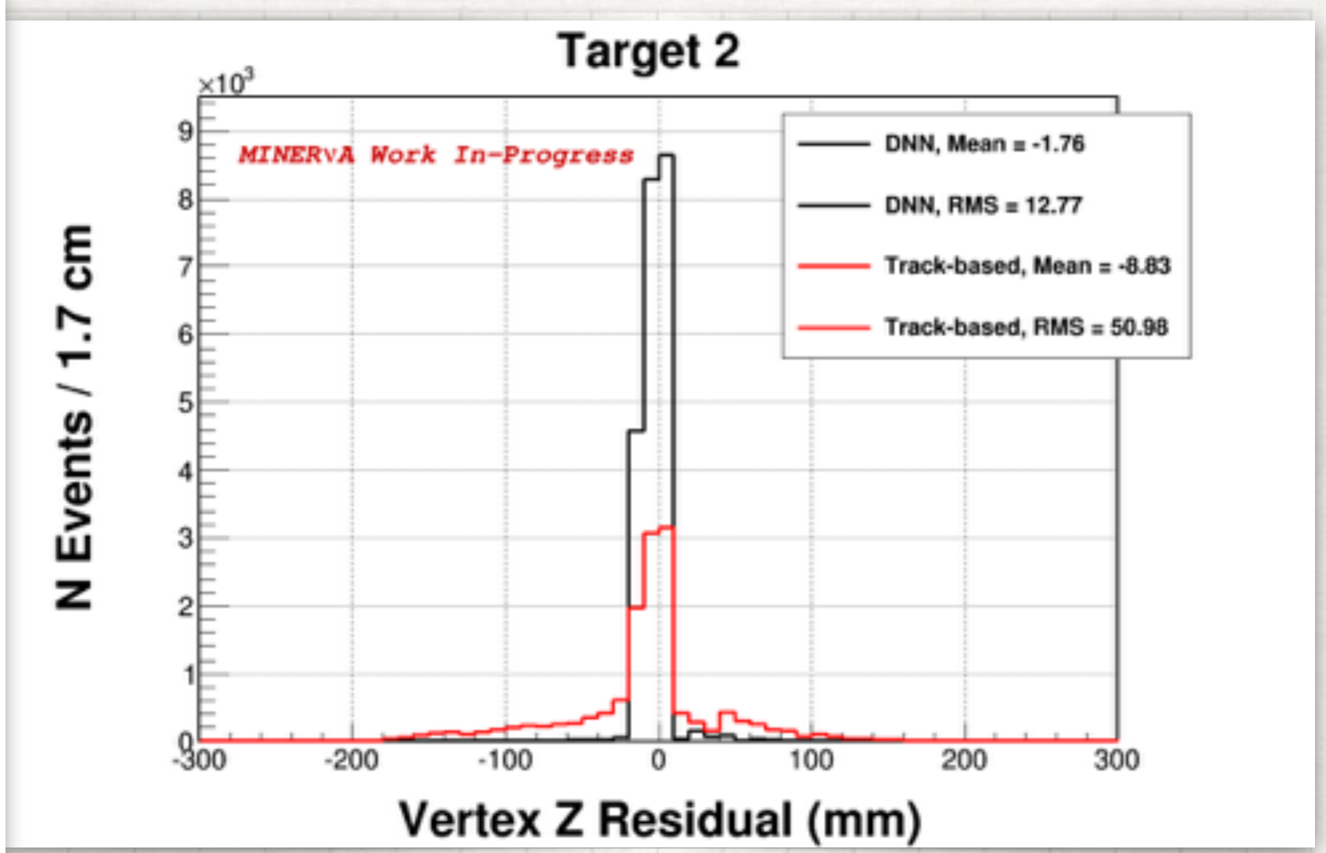


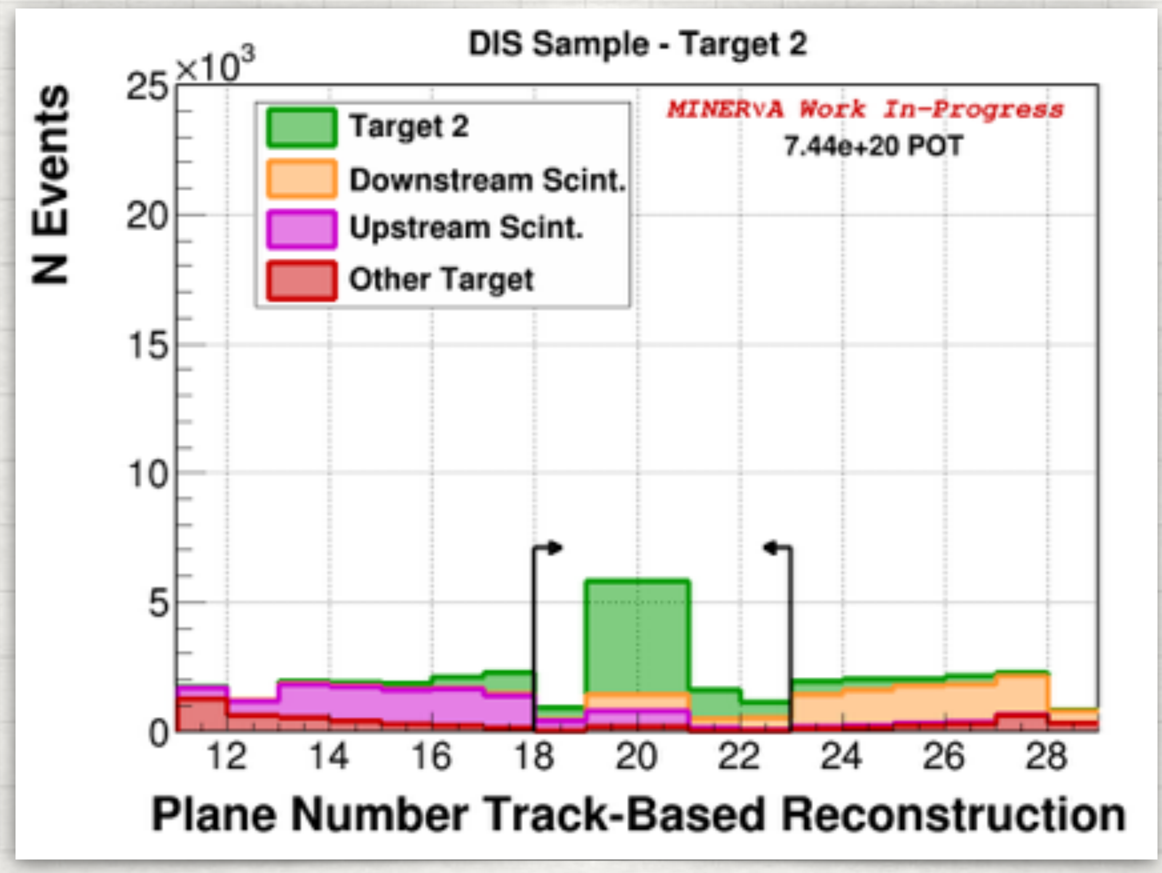
VERTEX FINDING RESULTS (SELECTED)

SEGMENT DCNN

Target	Track-Based Row Normalized Event Counts + stat error (%)	DNN Row Normalized Event Counts + stat error (%)	Improvement + stat error (%)
Upstream of Target 1	41.11±0.95	68.1±0.6	27±1.14
1	82.6±0.26	94.4±0.13	11.7±0.3
Between target 1 and 2	80.8±0.46	82.1±0.37	1.3±0.6
2	77.9±0.27	94.0±0.13	16.1±0.3
Between target 2 and 3	80.1±0.46	84.8±0.34	4.7±0.6
3	78±0.3	92.4±0.16	14.4±0.34







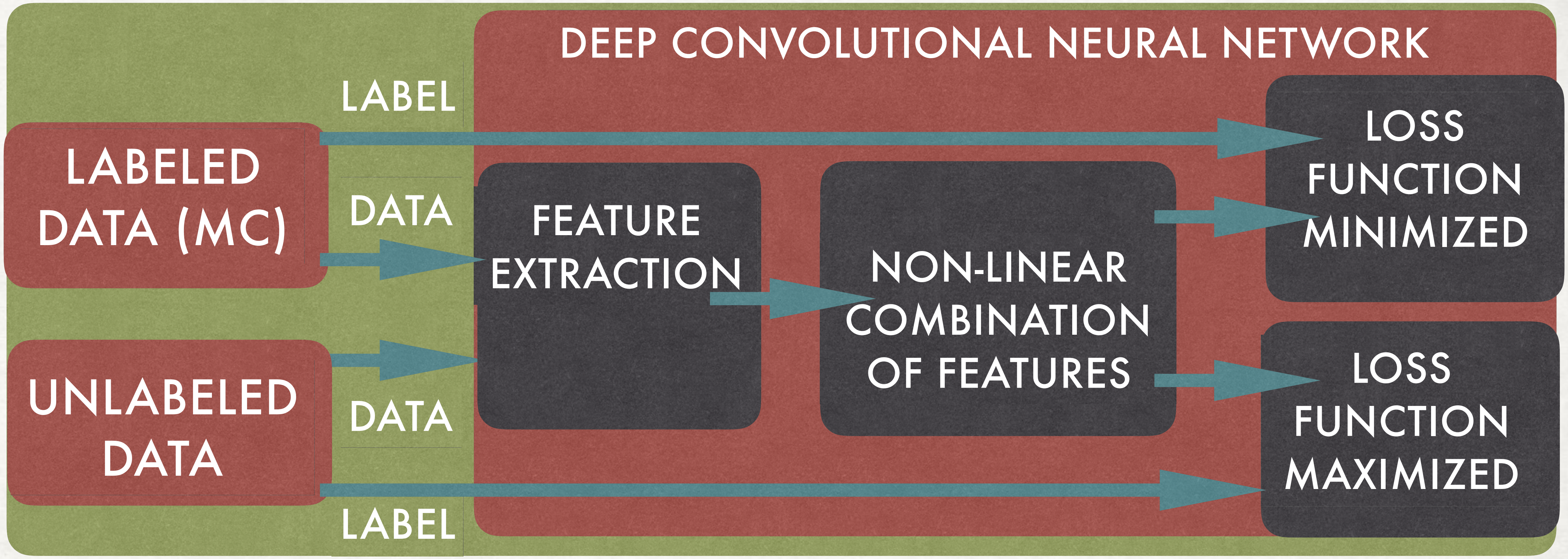
Here are results from the plane number classifier (67 planes).

Residual is true - center of plane for DNN and true - reconstructed z for track based reconstruction.

Regression was nonproductive for non-uniform/non-linear space studied.

MACHINE LEARNING

DOMAIN ADVERSARIAL TRAINING



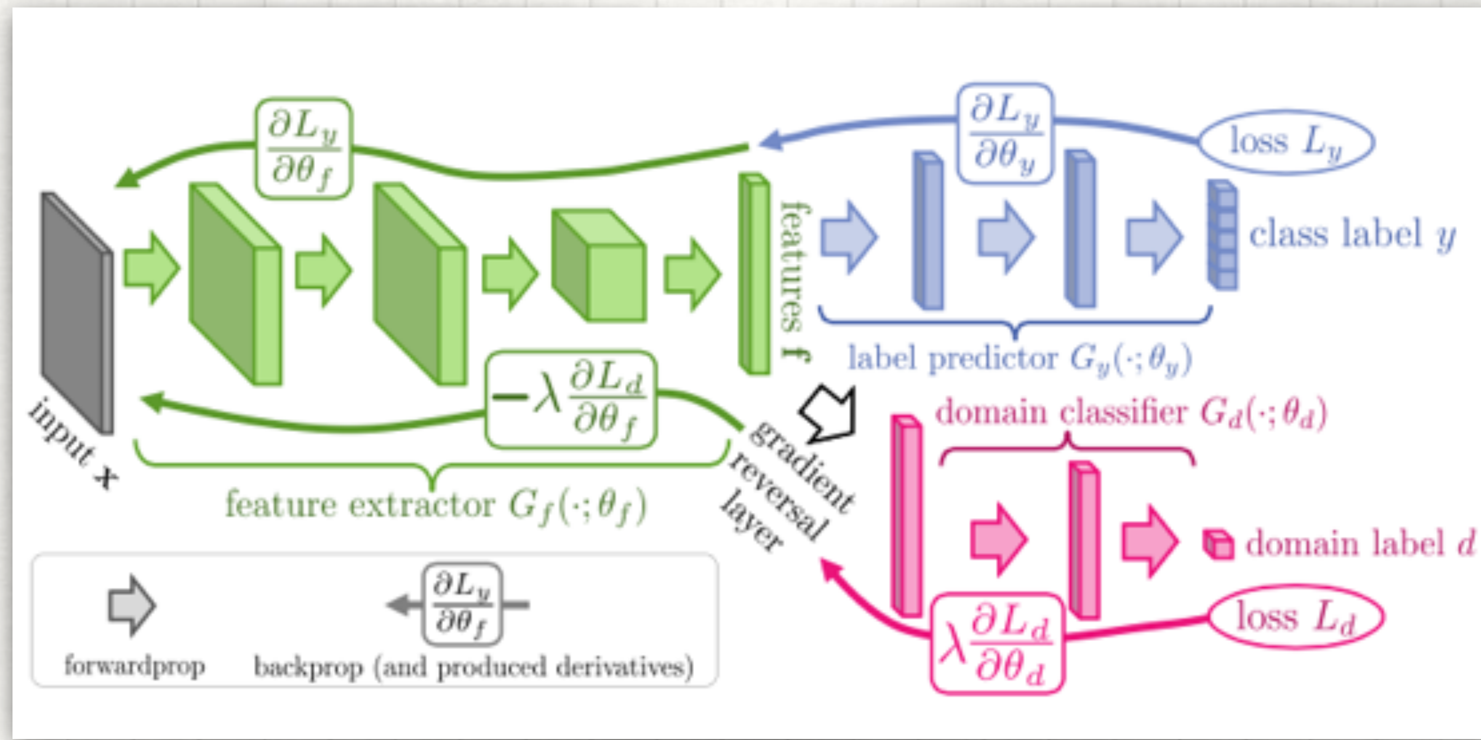
- In computer vision and pattern recognition a lack of labelled data is the problem, for us the problem is imperfect labeled data (simulation).

DOMAIN ADVERSARIAL TRAINING

DEEP NEURAL NETWORKS

- The training needs to be able to discriminate on the source domain but be indiscriminate between the domains.
- Training to extract and combine features is on the forward propagation, training to remove features which can be used to differentiate the domains on back propagation.
- The network develops an insensitivity to features that are present in one domain but not the other, and trains only on features that are common to both domains.

Combine simulation image and data image.



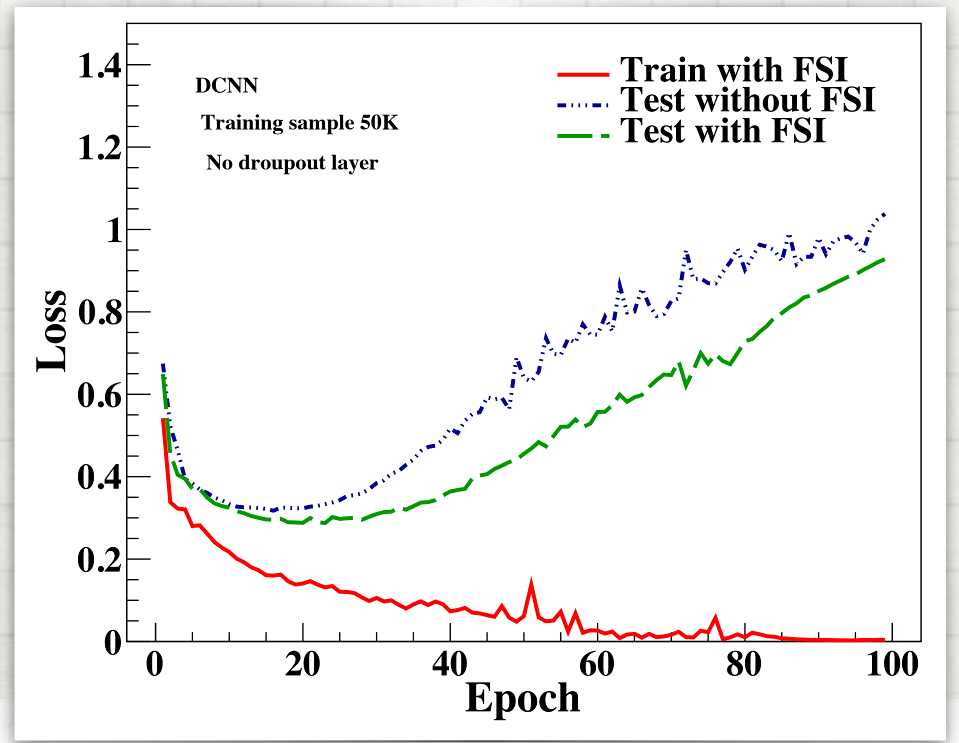
https://arxiv.org/abs/1505.07818

TESTS OF DOMAIN ADVERSARIAL TRAINING

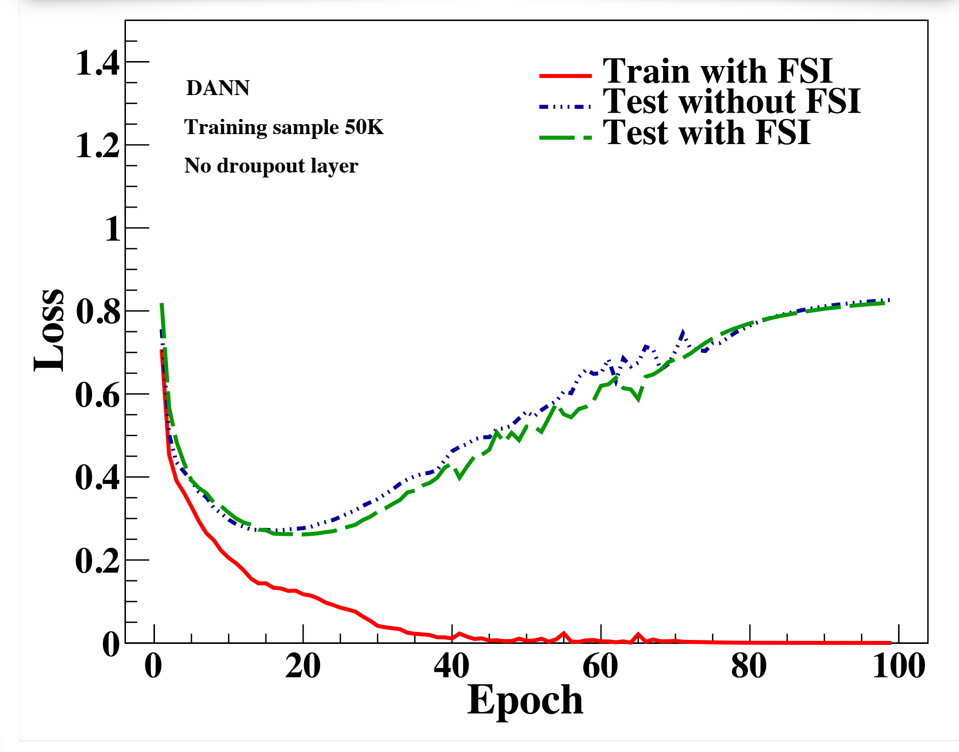
FINAL STATE INTERACTIONS (FSI)

- FSI is one of the central nuclear physics `corrections' that impact every measurement.
- We can see the effect of having different features between two domains by restricting our training samples, removing dropout layers and having different simulation as the target domain.
- For the NN with domain-adversarial training the loss increases at a slower rate and the behavior of the sample with both nuclear physics models (FSI on and off in GENIE) was approximately the same.

standard DNN



with domain-adversarial training



DISCUSSION

DEEP CONVOLUTIONAL NEURAL NETWORKS

- The selected problem, vertex finding in MINERvA in the ME flux, was selected to be one immune to most simulation problems (flux, nuclear model, etc) and for how clear it was for human scanners.
    - We will investigate systematics the traditional way by varying simulation (flux, nuclear model, etc).
    - We can calculate the uncertainty due to using this ML method by training the DCNN with different sets of simulations and observing the systematic error.
    - Very successful, effective increase of DIS statistics in targets of >30%.
- We have varied flux, nuclear model, W, etc when studying domain-adversarial training. Look for a paper to appear sometime in the next two months. We will continue to study domain-adversarial training in hadron multiplicity and semantic segmentation based particle identification.

BACKUP

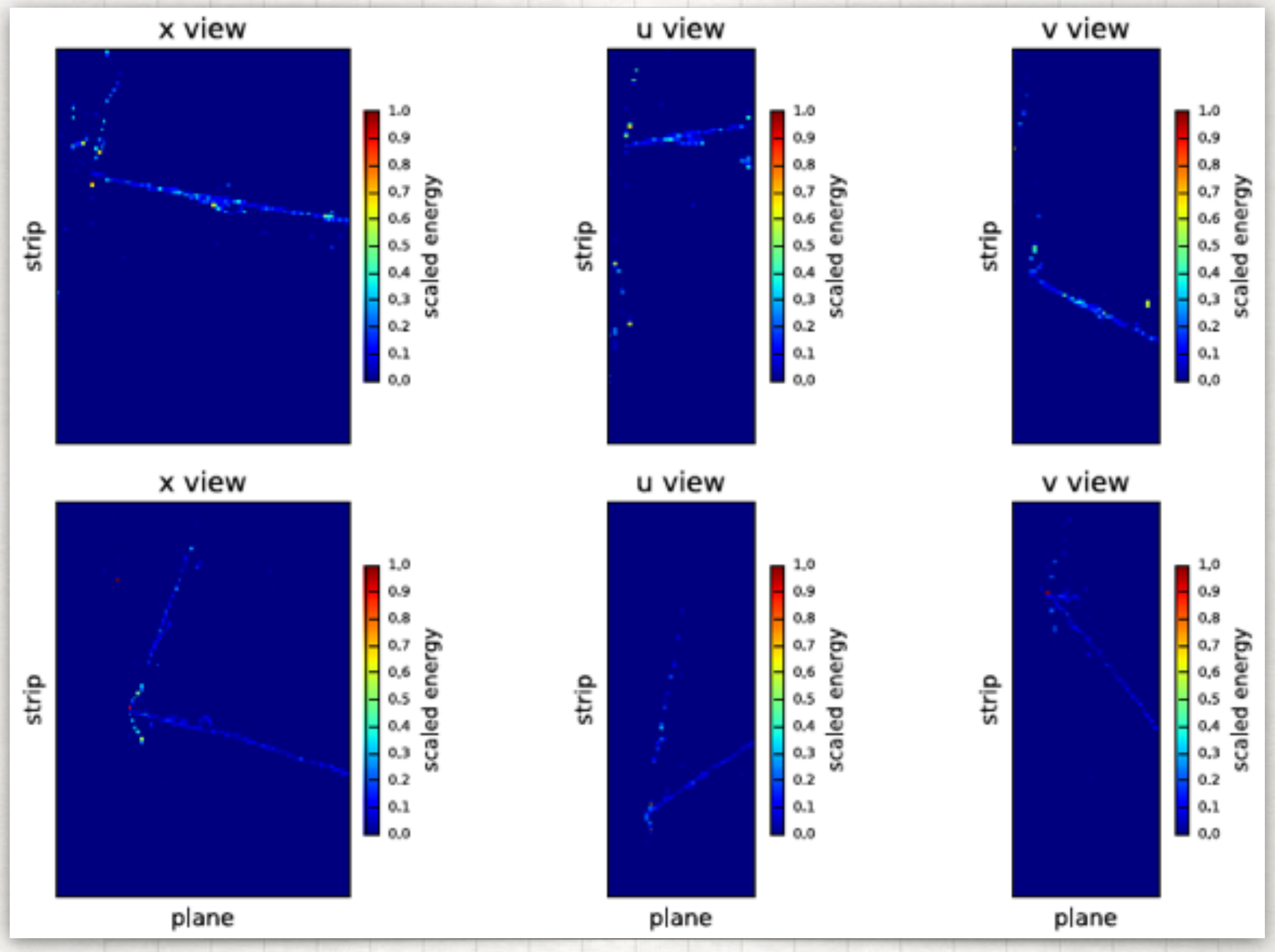
TYPES OF LAYERS

DONEC QUIS NUNC

- Convolutional layers: Normal neural layer - Made up of neurons with learnable weights. Convolution layers share weights across neurons.
- Pooling: as the number of feature maps grow, the complexity of the network explodes. Pooling reduces the "spatial size" or amount of parameters and computation in the network.
- Fully connected layer: Neurons in a fully connected layer have full connections to all activations in the previous layer.
- Dropout layer:Randomly drop connections between layers on each pass during training to eliminate co-adaptations in the network.
- Loss function :Loss function indicate the penalty for an incorrect prediction.

ENERGY IMAGES WITH NORMALIZED ENERGY WITHIN EACH EVENT.

MINERVA IMAGES



IMPROVEMENT OF THE VERTEX RECONSTRUCTION

TARGET 4 RESULTS