A Deep Learning Approach to Particle Identification for the AMS Electromagnetic Calorimeter

Raheem Hashmani[1], Emre Akbaş[2], Bilge Demirkoz[1], Zhili Weng[3]

[1]IVMER, Middle East Technical University (METU), 06800 Ankara, Turkey
[2]ImageLab, Middle East Technical University (METU), 06800 Ankara, Turkey
[3]Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA

About the Authors

METU - IVMER

Raheem Hashmani

B.S. in Physics, METU
Pursuing M.S. in Computer Engineering (CENG), METU

Prof. Dr. Bilge Demirköz

B.S., M.S. in Physics, MIT
PhD. In Physics, Oxford
Director, METU IVMER

METU - ImageLab

Prof. Dr. Emre Akbaş

B.S., M.S. in CENG, METU
PhD. in ECE, UIUC
Co-PI, METU ImageLab

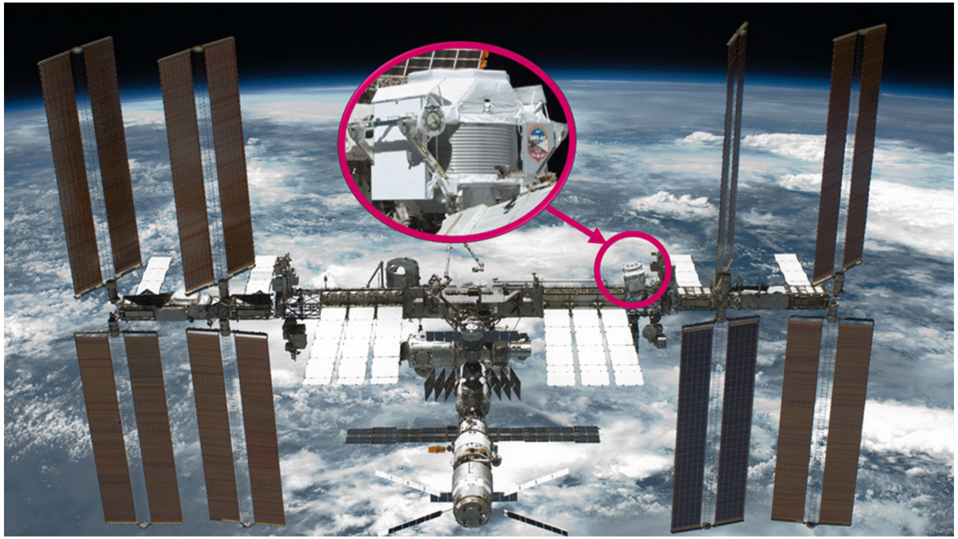
MIT - AMS

Dr. Zhili Weng

B.S., PhD. In Physics, Sun Yat-sen University
Principle Research Scientist, MIT

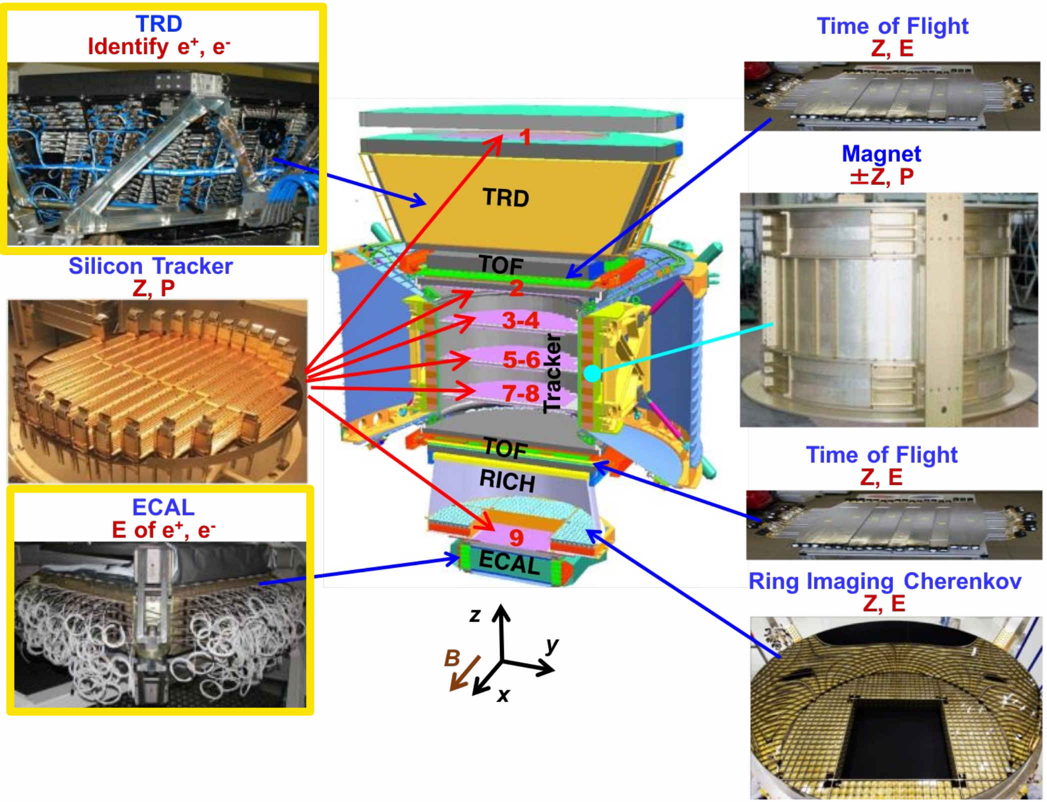
The Alpha Magnetic Spectrometer Experiment

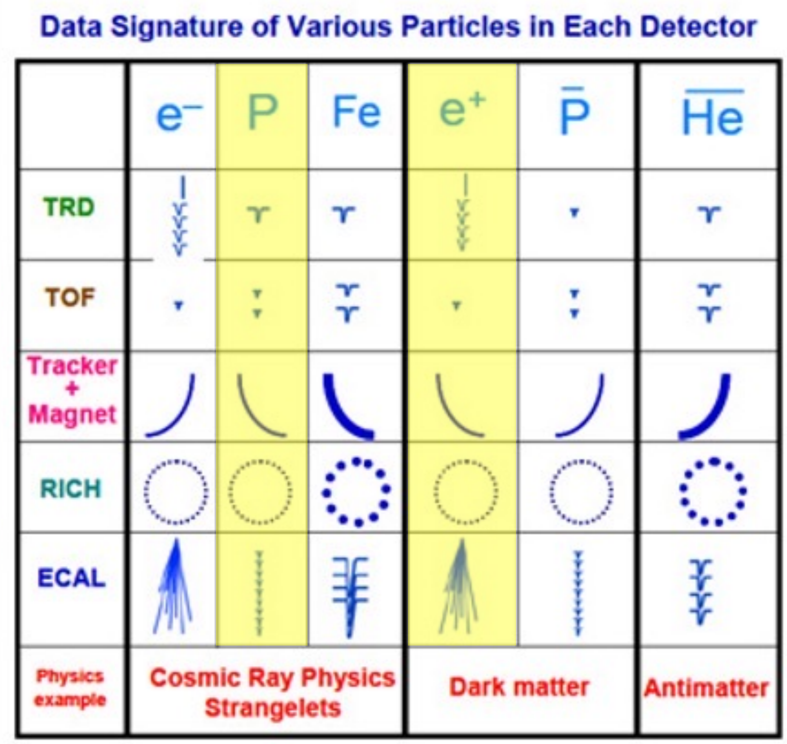
- General-purpose high-energy particle physics detector onboard the ISS [1].
- Installed on 19 May 2011
- Collected over 200 billion cosmic ray events
- Main objectives:
  - Searching for antimatter
  - Investigating dark matter
  - Analyzing cosmic rays



[1] S. Ting, Nuclear Physics B - Proceedings Supplements, Oct. 2013, doi: 10.1016/j.nuclphysbps.2013.09.028

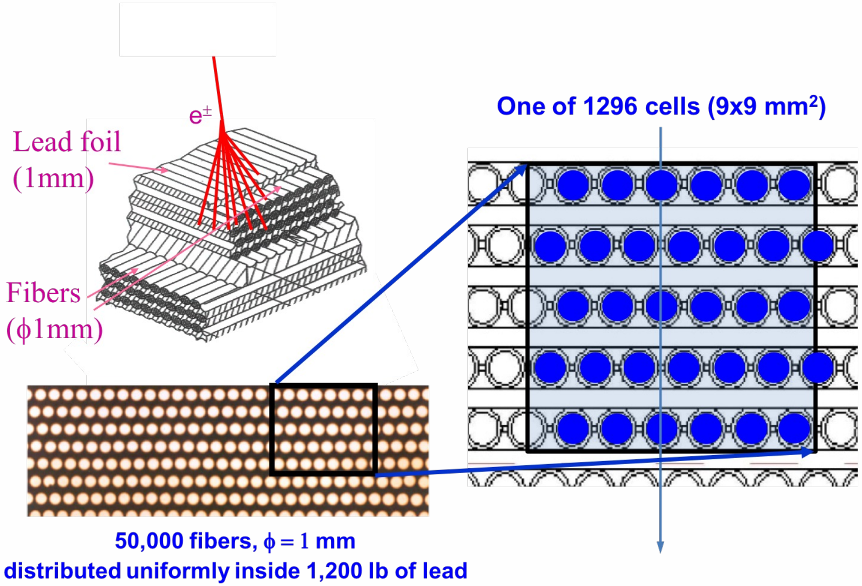
The Alpha Magnetic Spectrometer Experiment



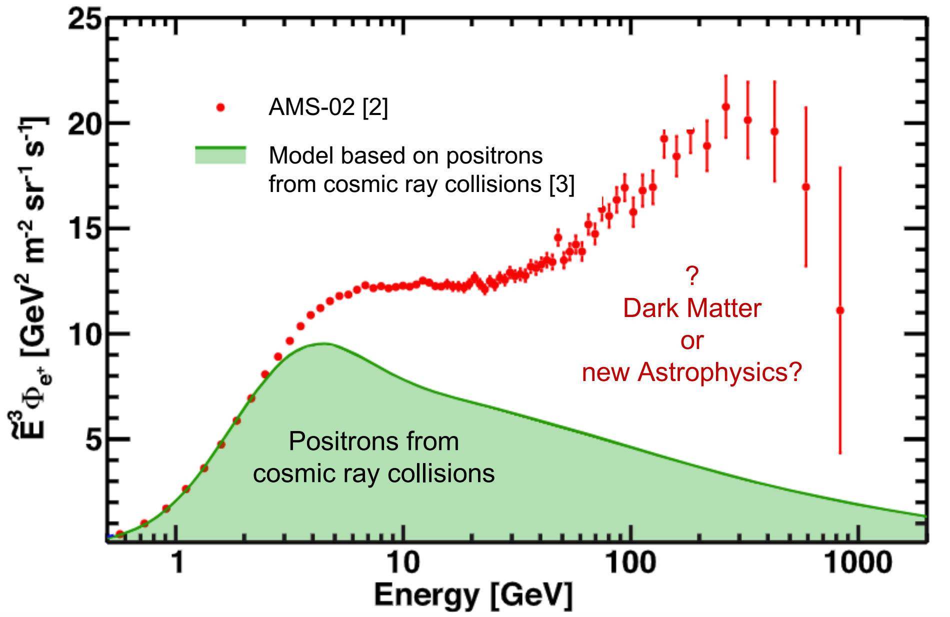


The AMS Electromagnetic Calorimeter

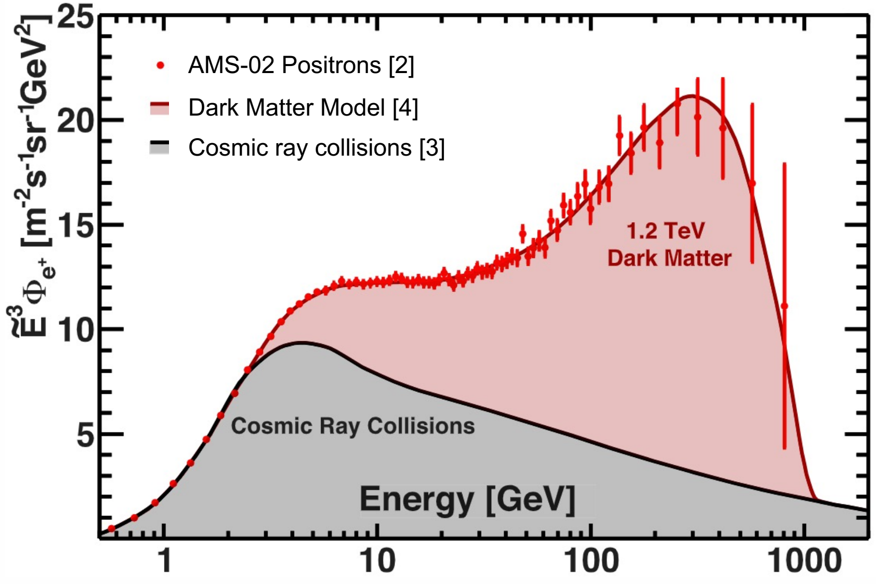
- Lead-scintillating fibres, running along one direction.
- Photomultiplier tubes at the end collect the generated light.
- Measures Energy Deposition
- 3D imaging capability
  - 648 x 648 x 166 mm^3
  - Depth of 17 radiation lengths
  - 18 cells for depth, 72 cells for the x/y axis.
- 18 layers, 10 for X axis, 8 for y axis



Physics Motivation for pure positron sample:



These results can not be explained by traditional cosmic ray models.



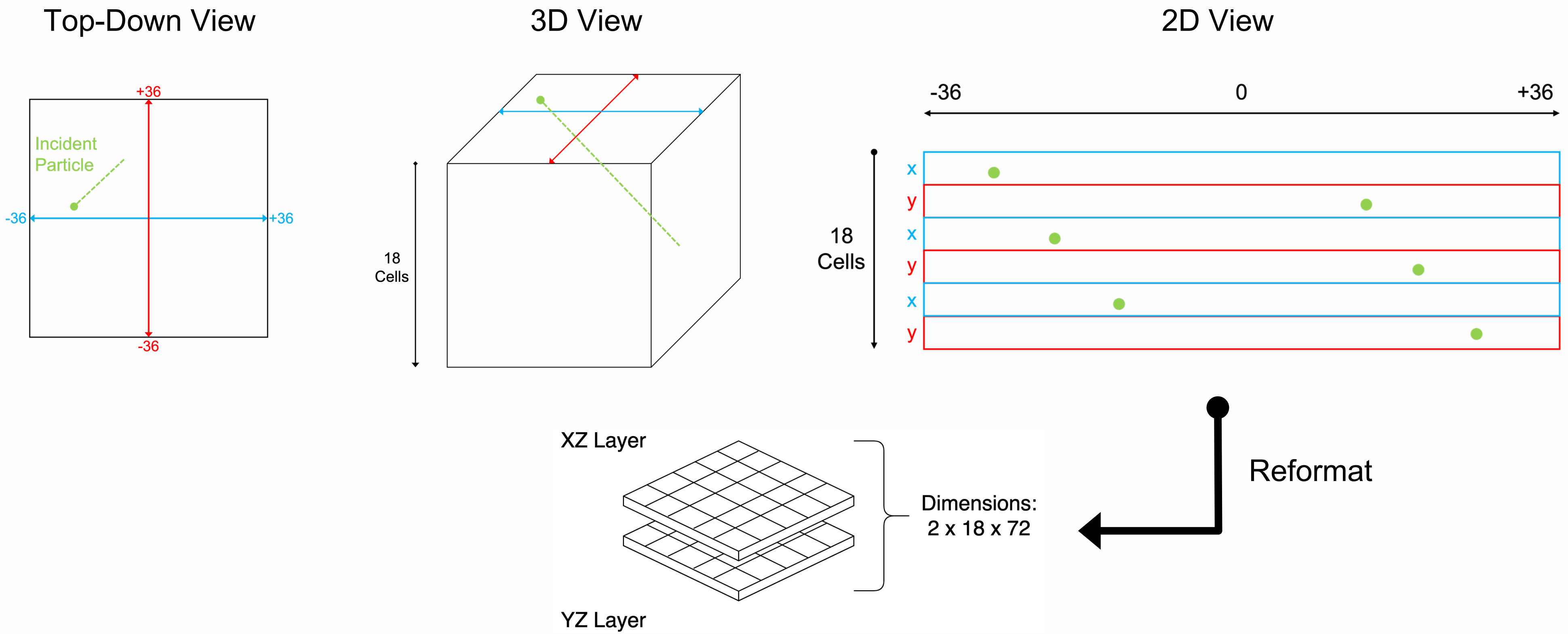
AMS Results are consistent with a Dark Matter Model.

[2] AMS Collaboration *et al.*, Phys. Rep., vol. 894, pp. 1–116, Feb. 2021, doi: 10.1016/J.PHYSREP.2020.09.003.
[3] R. Trotta *et al.*, Astrophys. J., vol. 729, no. 2, p. 106, Mar. 2011, doi: 10.1088/0004-637X/729/2/106.
[4] J. Kopp, Phys. Rev. D, vol. 88, no. 7, p. 076013, Oct. 2013, doi: 10.1103/PhysRevD.88.076013.

Overall Goal

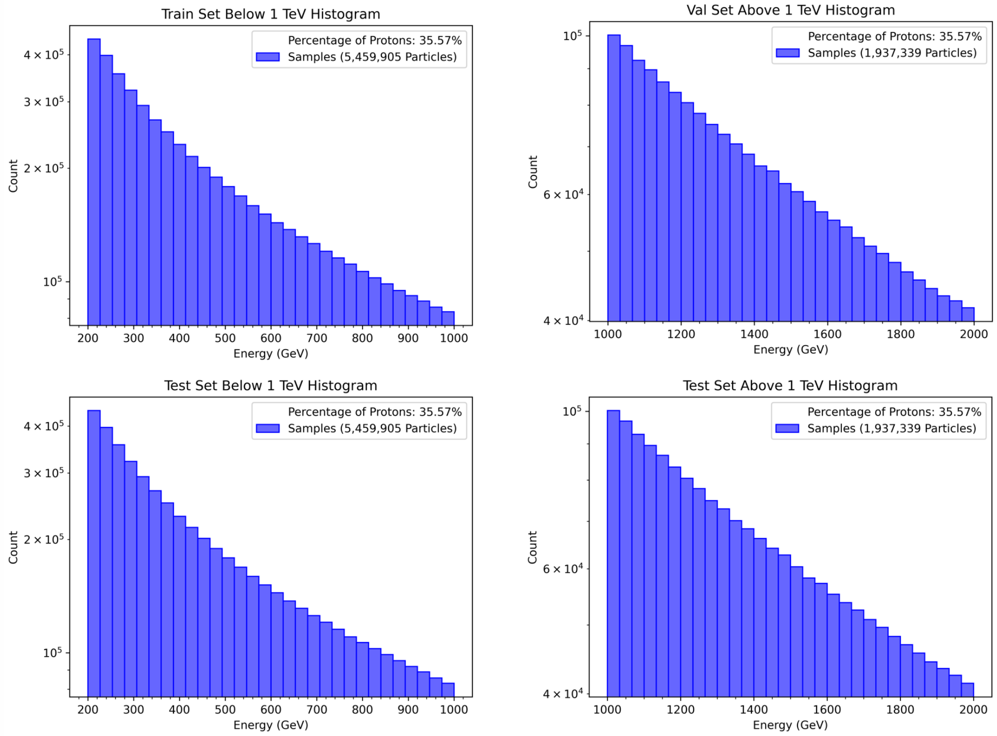
- Improving cosmic electrons/positrons identification over proton background at TeV energies.
  - To identify rare high energy positron events for analysis.
- Finding a deep learning model suitable for the task.
  - Fewer handcrafted features.
  - More efficient using GPUs.
  - Potential for scaling to other calorimeters.
- Creating a Model that Minimizes Dependency on Energy
  - After Monte Carlo, we will use TRD to sample pure protons and electrons from AMS data.
  - Limited to below 1 TeV with TRD.
  - If model is not overly dependent on energy, and performance is good on below 1 TeV AMS data, we can reasonably expect good results above 1 TeV on real data.

Data Preprocessing



Dataset

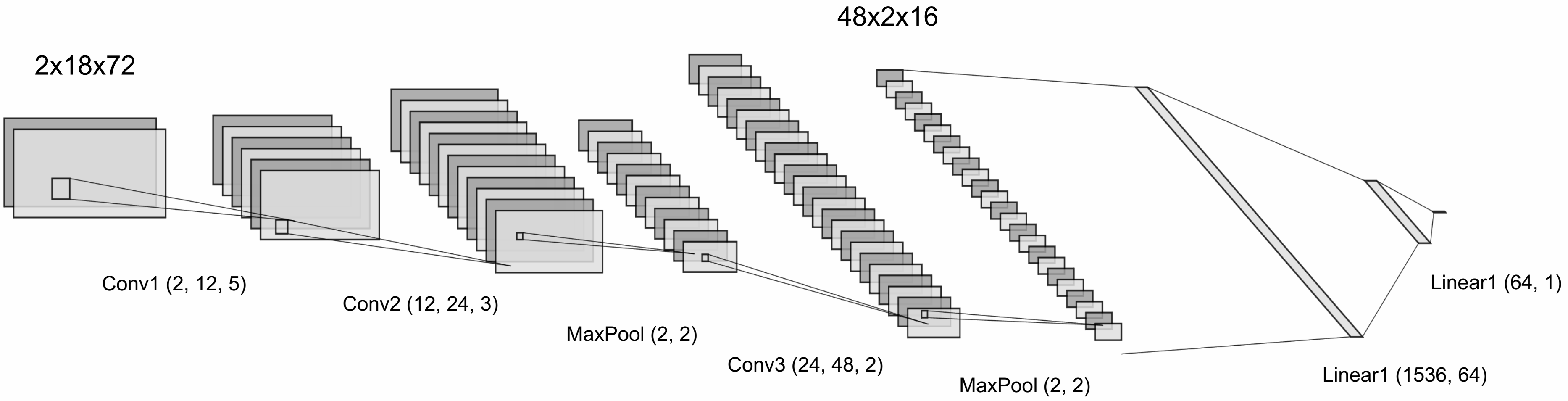
- AMS Generated Monte Carlo Data
- 200 – 1000 GeV:
  - Protons: 3.9 million images (~36% of total)
  - Electrons: 7.0 million images (~64% of total)
  - Split 50/50 into Train and Test Set
- 1000 – 2000 GeV:
  - Protons: 1.2 million (31% of total)
  - Electrons: 2.7 million images (69% of the total)
  - Split 50/50 into Val and Test Set



Models

- Trained many models (Linear Regression, SVM, Multilayer perceptron, etc.), this presentation will focus on 3:
- Convolutional Neural Network (CNN)
- Residual Neural Network (ResNet18)
- Convolutional Vision Transformer (CvT)

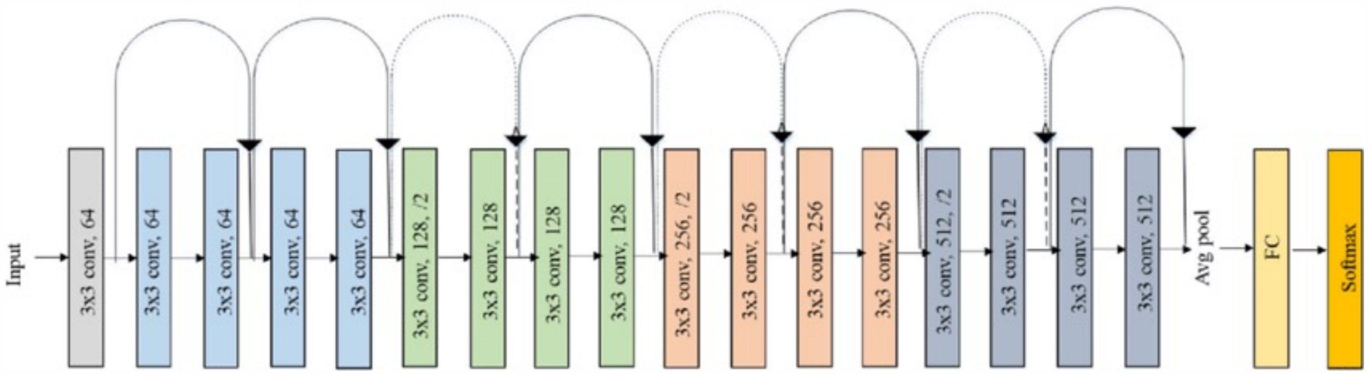
CNN



48x2x16
2x18x72
Conv1 (2, 12, 5)
Conv2 (12, 24, 3)
MaxPool (2, 2)
Conv3 (24, 48, 2)
MaxPool (2, 2)
Linear1 (64, 1)
Linear1 (1536, 64)

ResNet18

- 18 Blocks used - ResNet18 [5]
- Modified initial and final layer to take in 2x18x72 and output 1 neuron.
- Employ skip connections, which helps avoid vanishing gradients and reduces network degradation.



[5] K. He *et. al.*, “Deep Residual Learning for Image Recognition,” (CVPR), 2016.

CvT

- What is a Transformer?

CvT

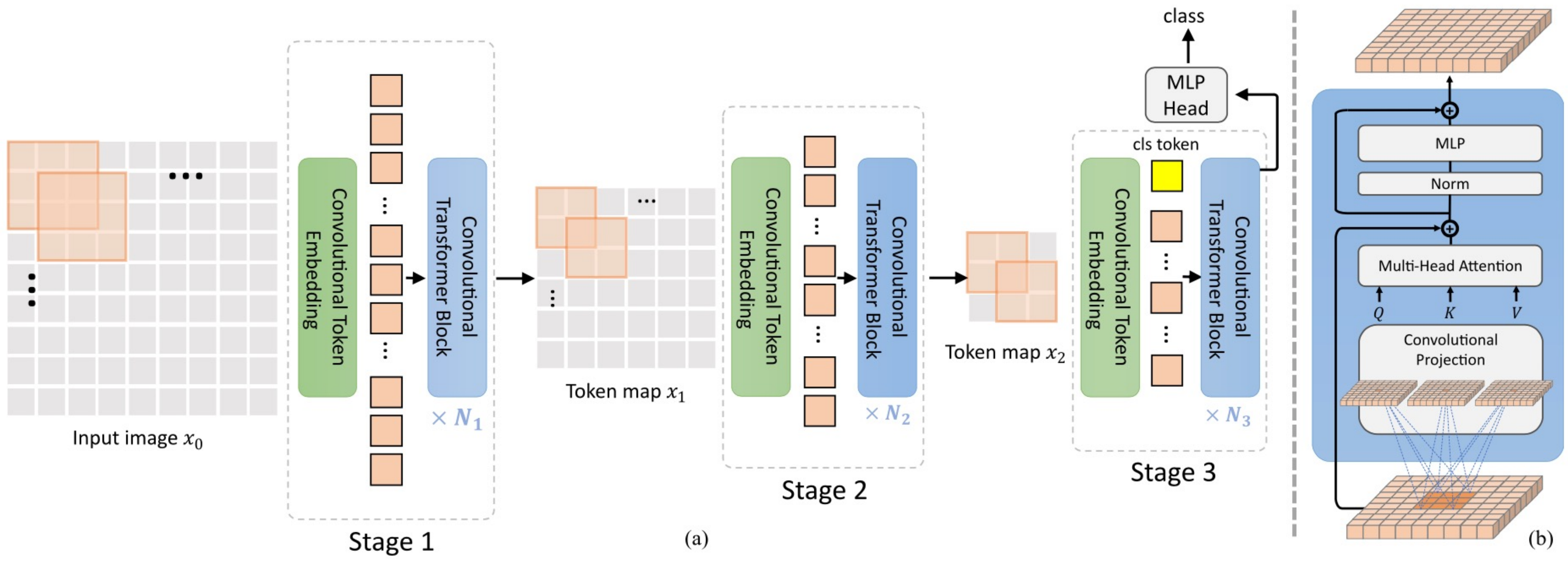
- What is a Transformer?

“Any architecture designed to process a connected set of units—such as the tokens in a sequence or the pixels in an image—where the only interaction between units is through self-attention.”

- What is Attention?
  - Method that dynamically highlights the most relevant parts of an input (key words in a sentence, key parts of an image).
- Commonly used in Recurrent Neural Networks, it was discovered that they are effective on their own for NLP [6].
- Found to be effective for image classification as well e.g., Vision Transformers [7]

[6] A. Vaswani et al., “Attention is All you Need,” Adv. Neural Inf. Process. Syst., vol. 30, 2017.
[7] A. Dosovitskiy *et al.*, “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale,” Oct. 2020

Convolutional Vision Transformer (CvT)

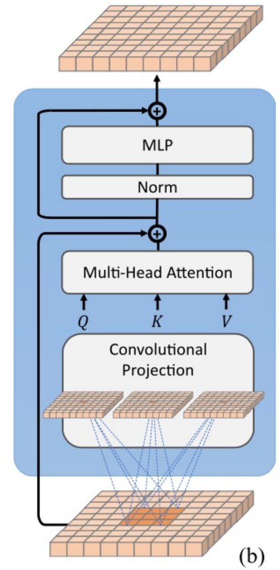


The pipeline of the CvT architecture [8]. (a) Overall architecture, showing the hierarchical multi-stage structure facilitated by the Convolutional Token Embedding layer. (b) Details of the Convolutional Transformer Block, which contains the convolution projection as the first layer.

[8] H. Wu *et al.*, Proc. IEEE Int. Conf. Comput. Vis., pp. 22–31, 2021, doi: 10.1109/ICCV48922.2021.00009.

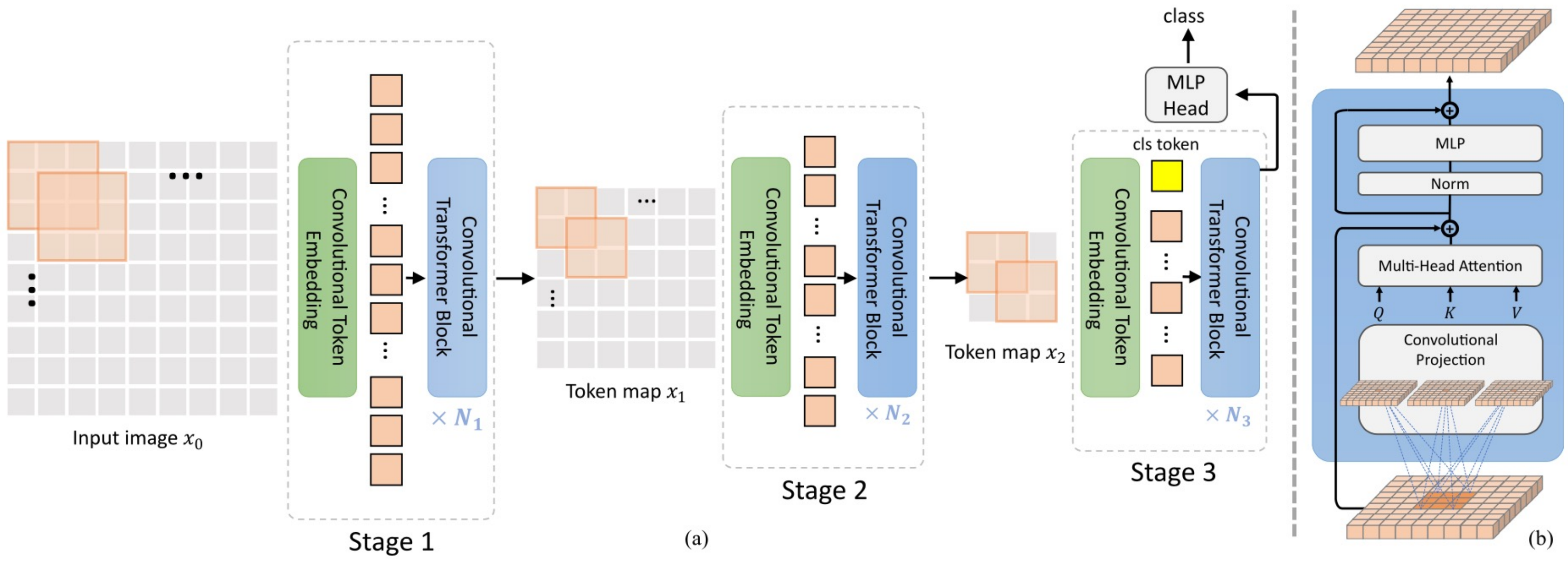
Convolutional Transformer Block

$$\mathrm{Attention}(Q, K, V) = \mathrm{softmax}(\frac{QK^T}{\sqrt{d_k}})V$$



(b)

Convolutional Vision Transformer (CvT)

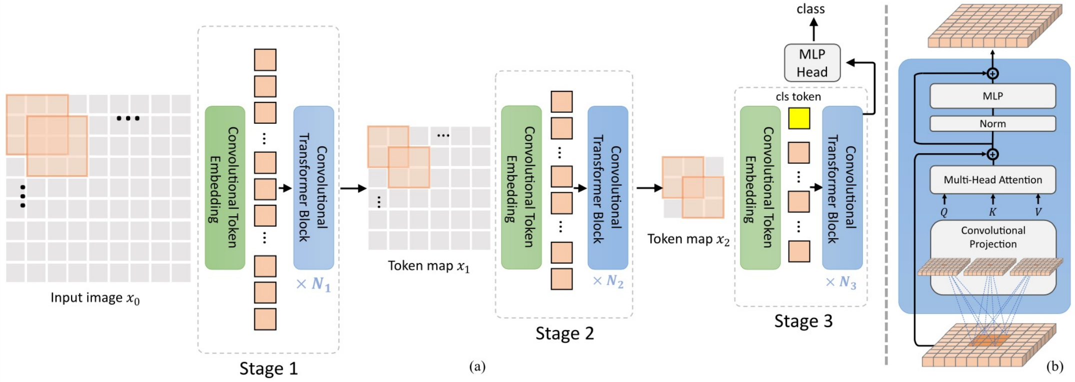


The pipeline of the CvT architecture [8]. (a) Overall architecture, showing the hierarchical multi-stage structure facilitated by the Convolutional Token Embedding layer. (b) Details of the Convolutional Transformer Block, which contains the convolution projection as the first layer.

[8] H. Wu *et al.*, Proc. IEEE Int. Conf. Comput. Vis., pp. 22–31, 2021, doi: 10.1109/ICCV48922.2021.00009.

CvT

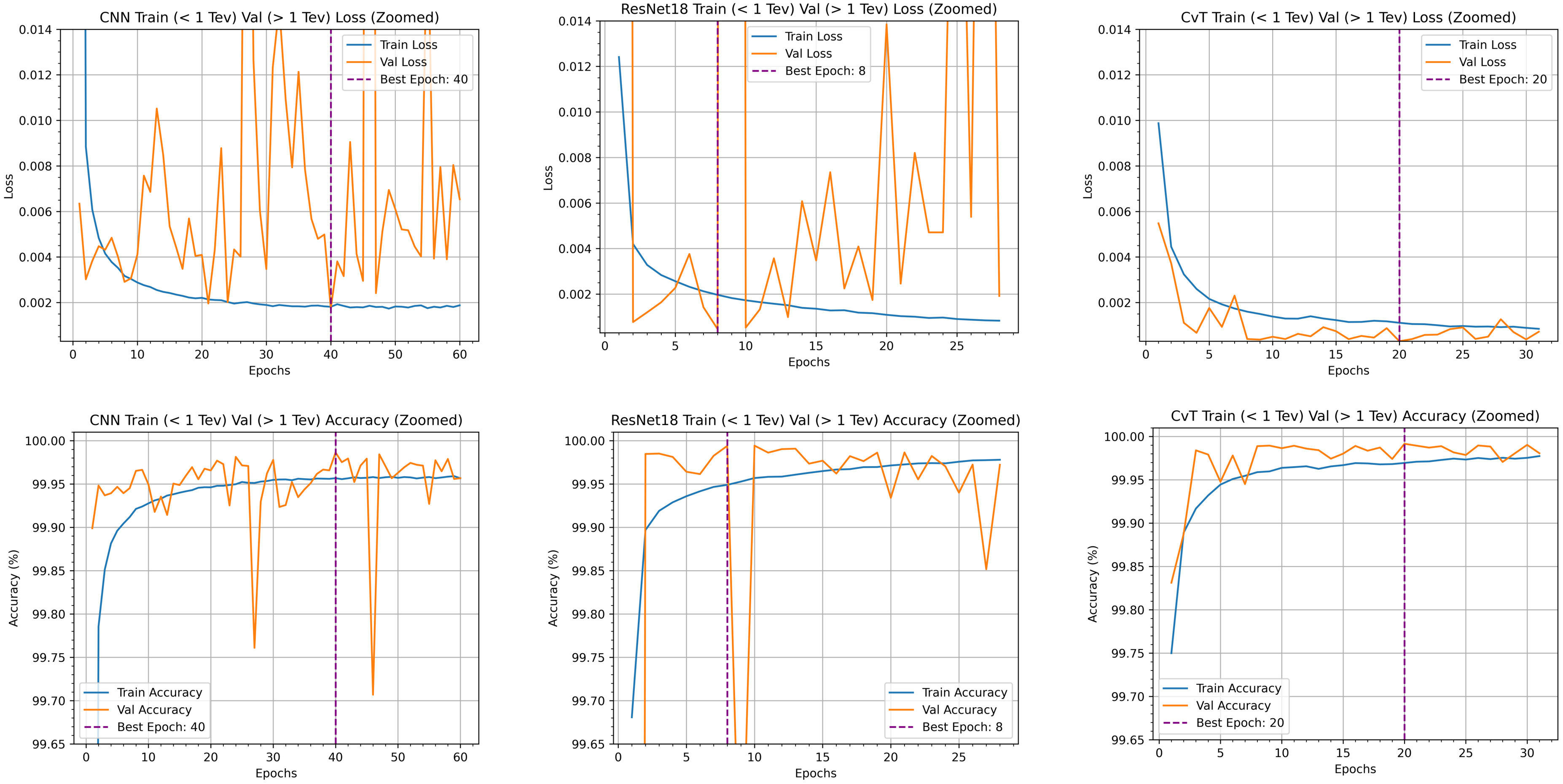
- Combines the benefits of CNNs (scale, shift, rotation, and distortion invariance) with Transformers (better generalization, better focus on key areas, global context).
- Important for ECAL showers ➔ shower “images” show different deposition shapes depending on the energy and type of particle, angle of incidence, and point of entry.
- Our implementation slightly modified ➔ uses 4 Transformer blocks (Stage 3 having 2 blocks), and 1, 3, and 6 attention heads for each stage, respectively.



Overall Hyperparameters

Hyperparameter	Value
Batch Size	128
No. of Workers	4
Loss Function	Binary Cross Entropy with Logits, Weighted
Activation Function after Last Layer	Sigmoid
Optimizer	Adam
Learning Rate for Optimizer	1.0e-4
Epochs:	Early stopping, with patience of 20.

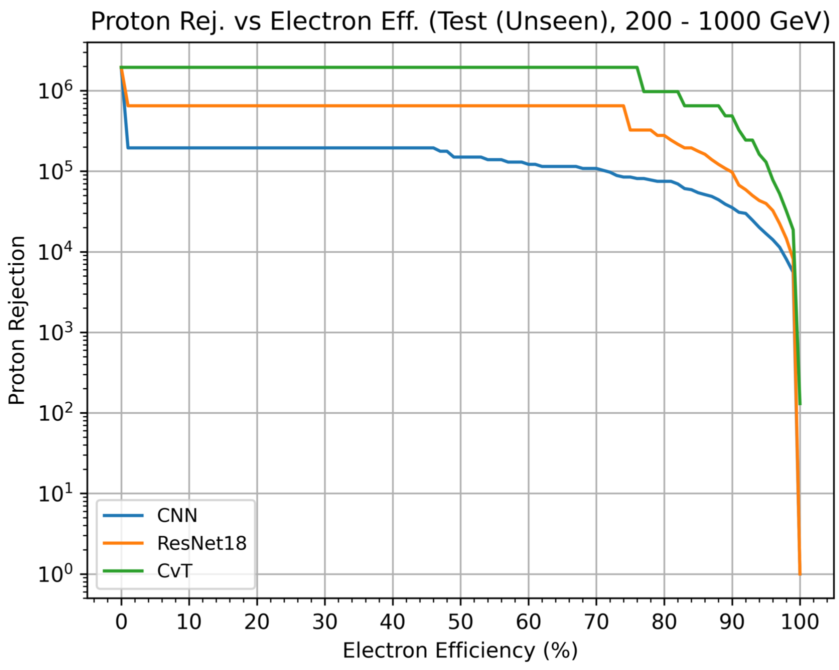
Training Results (Scaled to the same y axis)



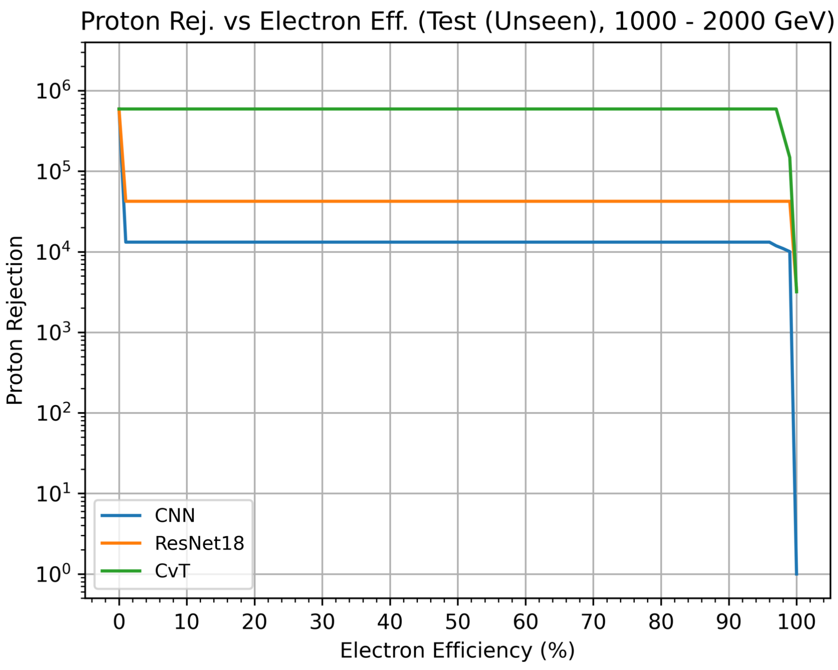
Evaluating Proton Rejection vs. Electron Efficiency

- Test set split into protons and electrons. Calculate the probability threshold, X, above which the electrons-only set gives a specific electron accuracy (efficiency), A_e.
  - E.g., given A_e=90%, we get X = 0.82. i.e., if we classify particles above 0.82 probability as electrons, we get an efficiency of 90%.
- We then take our threshold, X, for a particular A_e and determine the proton rejection on the proton-only test set:
  - $\text{Proton Rejection} = \dfrac{\text{Total number of protons}}{\text{Number of proton misidentified}}$
  - Where protons misidentified are particle that are known to be protons, getting a probability higher than X.

Trained on 200-1000 GeV, Evaluated on 1000 – 2000 GeV, Tested on Unseen Data of Both



At 90% e⁻ efficiency, the CvT outperforms the ResNet and CNN models by a factor of 5 and 14.



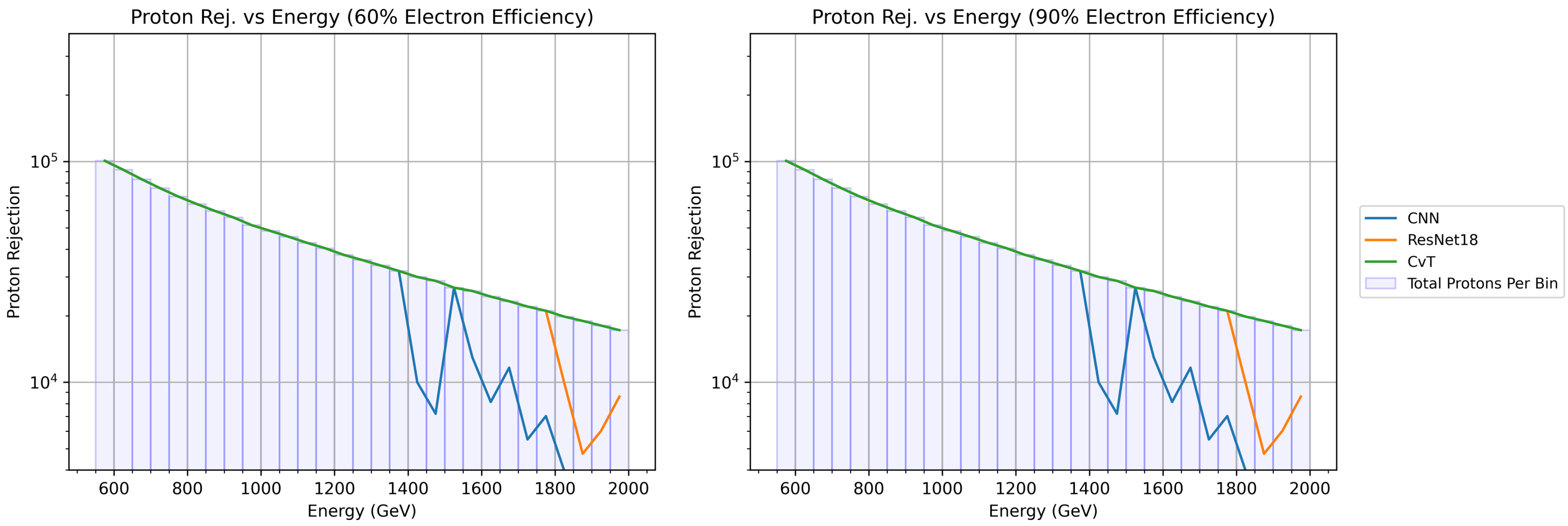
At 90% e⁻ efficiency, the CvT outperforms the ResNet and CNN models by a factor of 14 and 45

Evaluating Proton Rejection vs. Energy

- Like before, but also splitting the proton/electron test sets by energy bins of width 50 GeV.
- To show which energy intervals our model performs poorly at.
- Preliminary study on the energy dependence of the models.

Evaluating Proton Rejection vs. Energy

Trained: 200-1000 GeV. **Evaluated** at each epoch: 1000 – 2000 GeV



Conclusion

- Evaluated calorimeter particle identification potential using deep learning.
- Took as an input all the cells in the ECAL for vision-based models.
- Trained and tested 3 models, the CNN, ResNet18, and CvT.
- Found the CvT model to significantly outperform the CNN and ResNet 18 models for shower classification. Signals potential for CvT to be effective in other calorimeter experiments as well.
- Further study is ongoing to understand the energy dependence.

Outlook

- Test on real data sample of pure electrons and protons at low (less than 500 GeV) energy selected by other AMS subdetectors.
    - TRD provides independent measurement of electrons/positrons, but only below 1 TeV.
    - If our models have minimal dependency on energy, good results on below 1 TeV give confidence in performance above 1 TeV.
- Determining minimum amount of data required to effectively train the ECAL CvT model.
- Additional preprocessing to make the CvT more “physics-aware”. Stay tuned.

Thank You Very Much!

Any Question?