UNIVERSITY OF CAMBRIDGE

Neutrino Event Classification: CNNs and Transfer Learning

Leigh Whitehead

University of Cambridge

IPA-ML 2023, ETH Zürich, March 2023

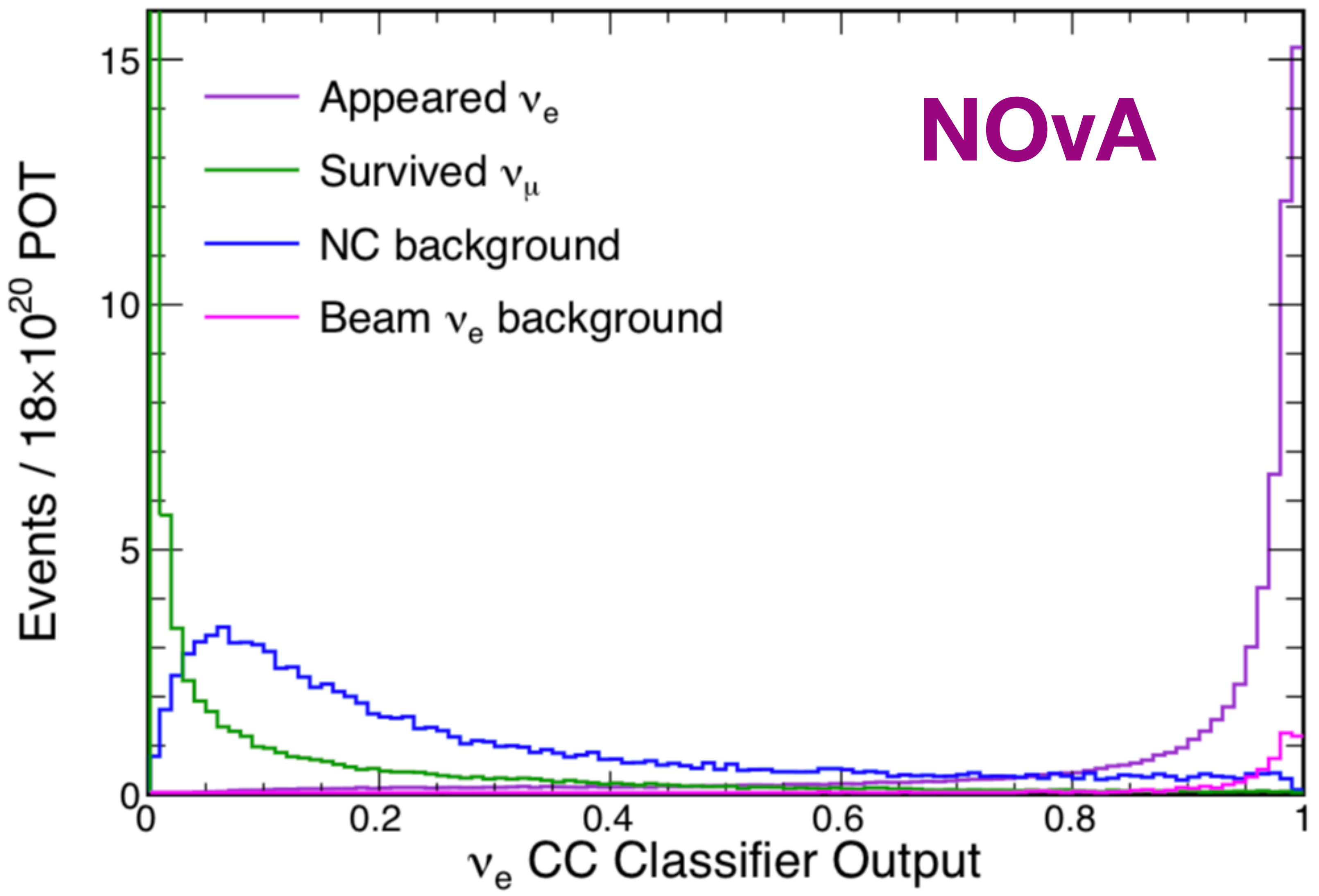
Introduction

- I will briefly discuss the use of CNNs in neutrino event classification
  - With a focus on the DUNE algorithm
  - Well-established technique

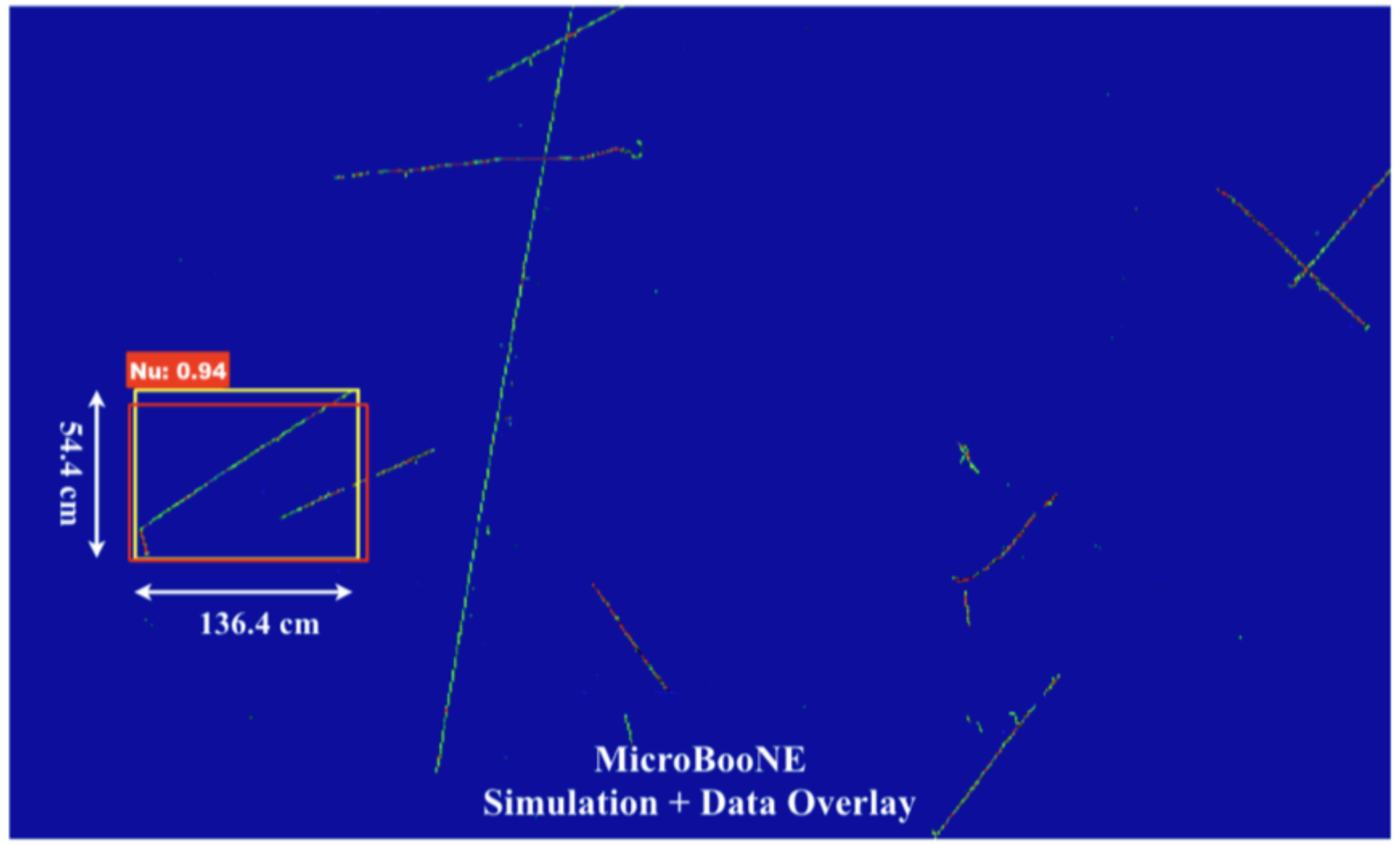
- I will move on to discussing a recent study on transfer learning
  - I hope this will have applications outside of neutrino physics

CNNs in neutrino physics

- The NOvA experiment was the first to use a CNN[1]
  - Used for event classification
  - 40% increase in efficiency with no loss of purity for their main CC ν_e analysis



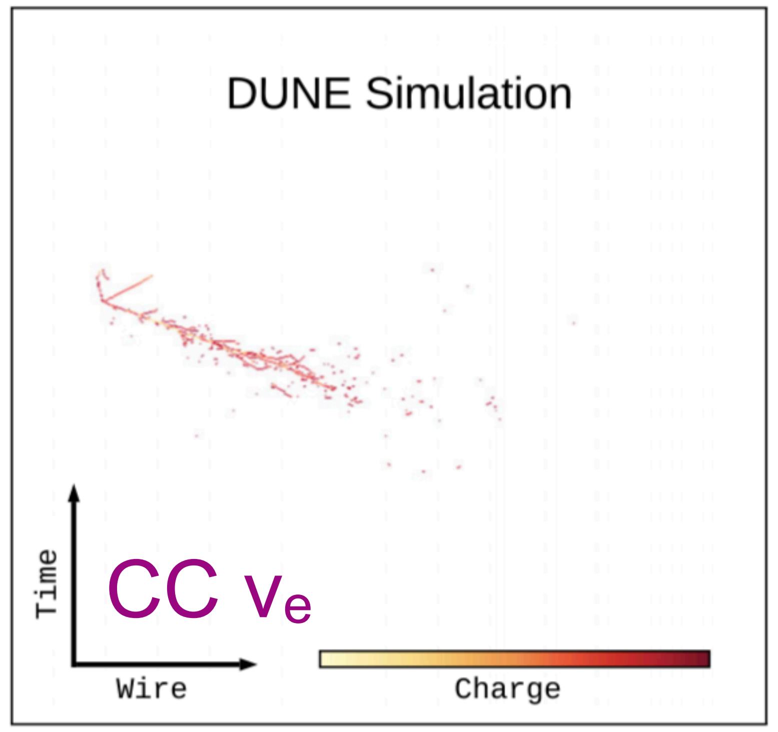
- MicroBooNE: first LArTPC experiment to use a CNN[2]
  - Used for region of interest finding and event classification

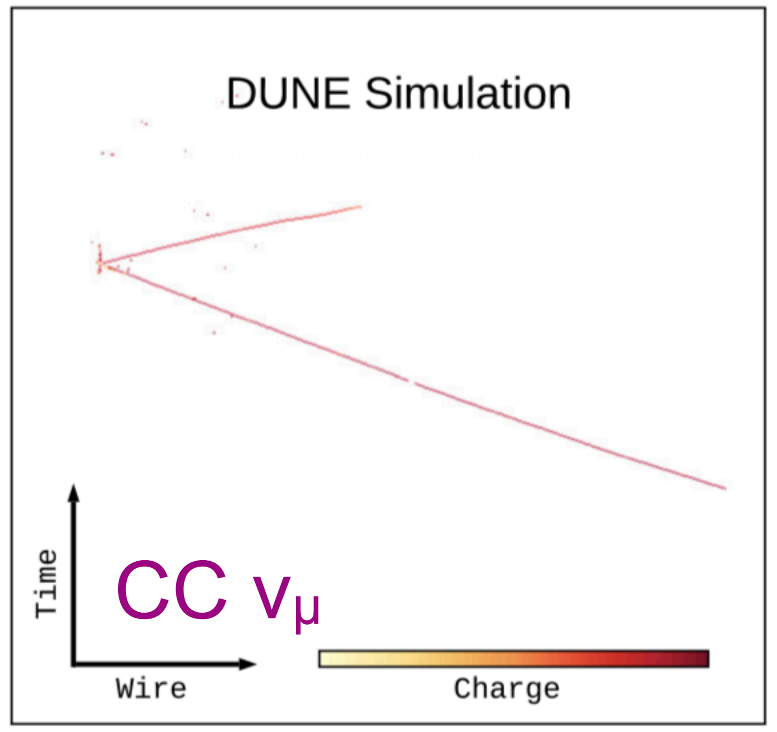


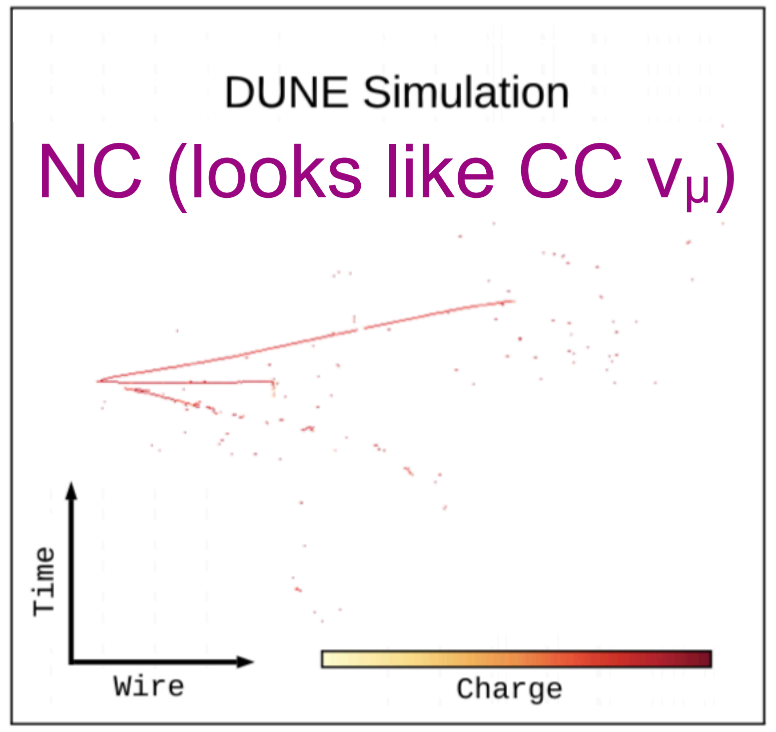
[1] A. Aurisano, et al., *A convolutional neural network neutrino event classifier,* Journal of Instrumentation 11 (2016) 09, P09001
[2] MicroBooNE Collaboration, *Convolutional Neural Networks Applied to Neutrino Events in a Liquid Argon Time Projection Chamber*, JINST 12 (2017) 03, P03011

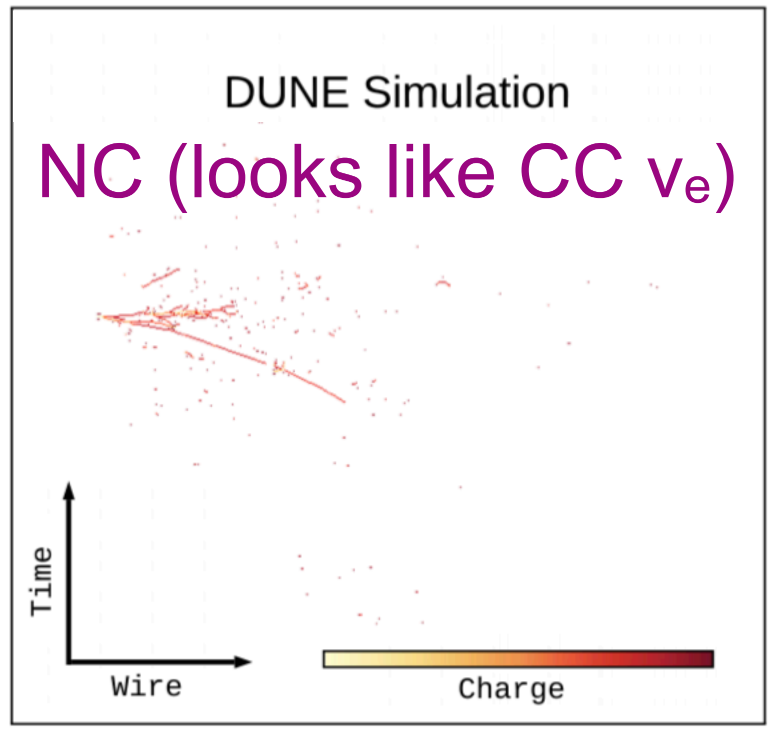
DUNE

- The Deep Underground Neutrino Experiment (DUNE) is a next generation neutrino oscillation experiment
  - Uses liquid argon time projection chamber technology (LArTPC)
  - Three 2D projections of each interaction sharing one common coordinate
- DUNE CVN[1] aims to classify events as CC ν_μ, CC ν_e, CC ν_τ, and NC
  - CC ν_τ are rare and hard to classify, so I won't discuss them further





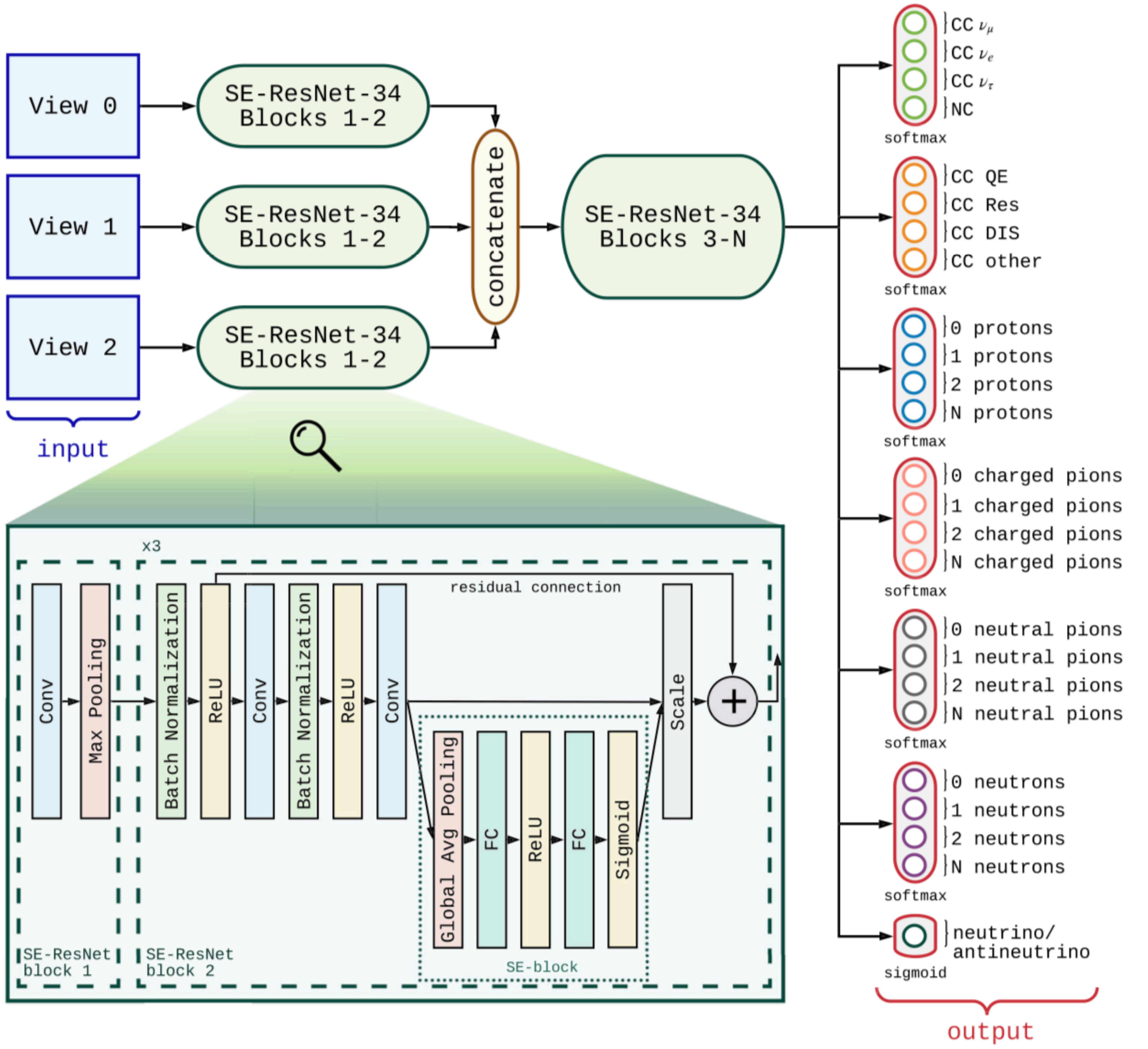




[1] DUNE Collaboration, *Neutrino interaction classification with a convolutional neural network in the DUNE far detector,* Phys.Rev.D 102 (2020) 9, 092003.
[2] S. Alonso Monsalve, *Novel usage of deep learning and high-performance computing in long-baseline neutrino oscillation experiments*, PhD Thesis, Universidad Carlos III Madrid (2021)

DUNE CVN

- Architecture based on SE-ResNet-34[1,2]
- Inputs processed separately for the first few blocks and then merged
- Main output is the flavour classifier
  - The top one shown in the figure
- Other particle counting outputs will be further studied in the future
- Trained on over 3 million events



[1] K. He, X. Zhang, S. Ren, and J. Sun, *Deep Residual Learning for Image Recognition*, 1512.03385; K. He, X. Zhang, S. Ren, and J. Sun, *Identity Mappings in Deep Residual Networks*, 1603.05027
[2] J. Hu, L. Shen, and G. Sun, *Squeeze-and-Excitation Networks*, 1709.01507

DUNE CVN

- See very good signal background separation

Arrows show events selected for the CC ν_e appearance sample

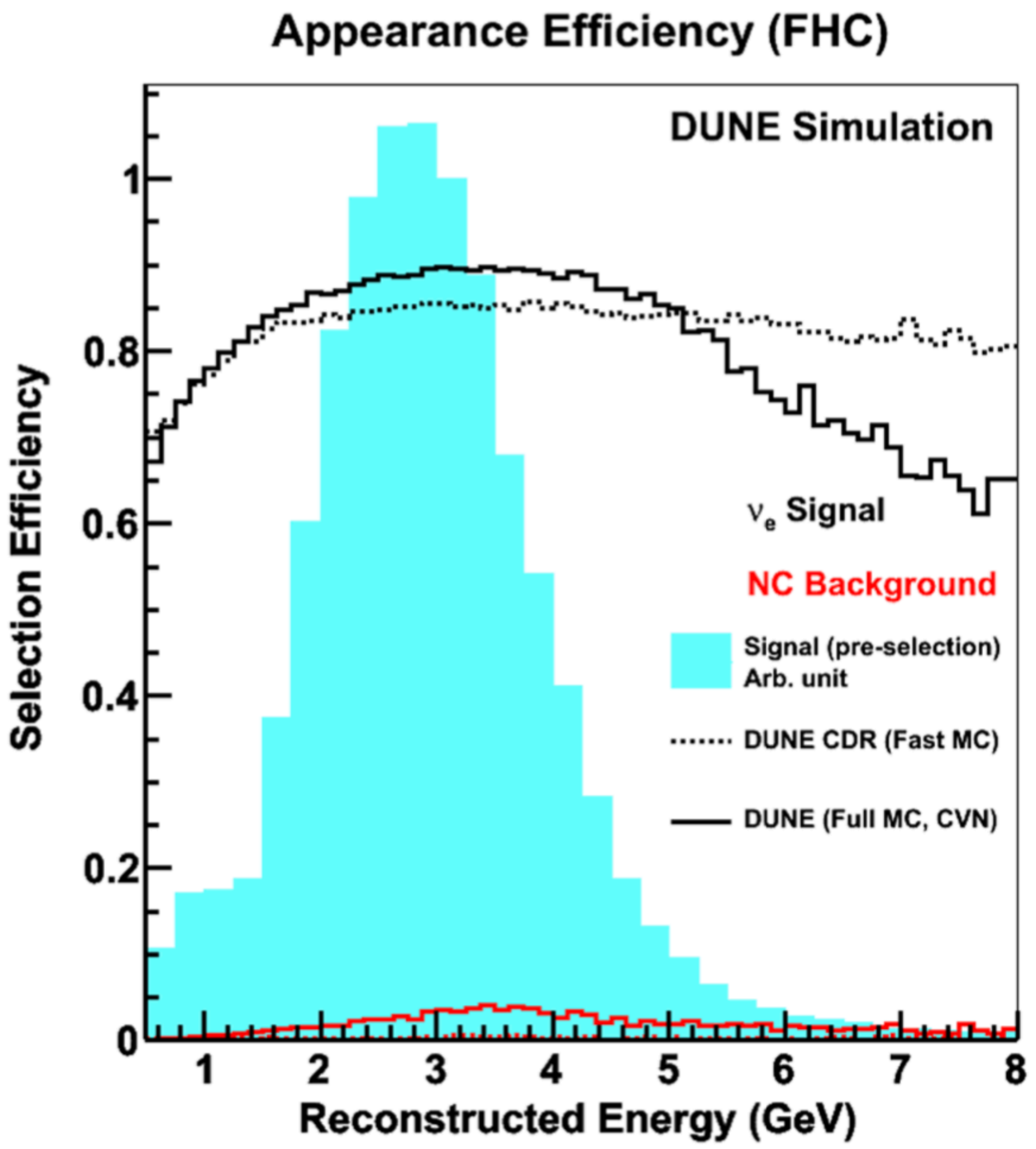
Arrows show events selected for the CC ν_μ disappearance sample

[1] DUNE Collaboration, *Neutrino interaction classification with a convolutional neural network in the DUNE far detector*, Phys.Rev.D 102 (2020) 9, 092003.

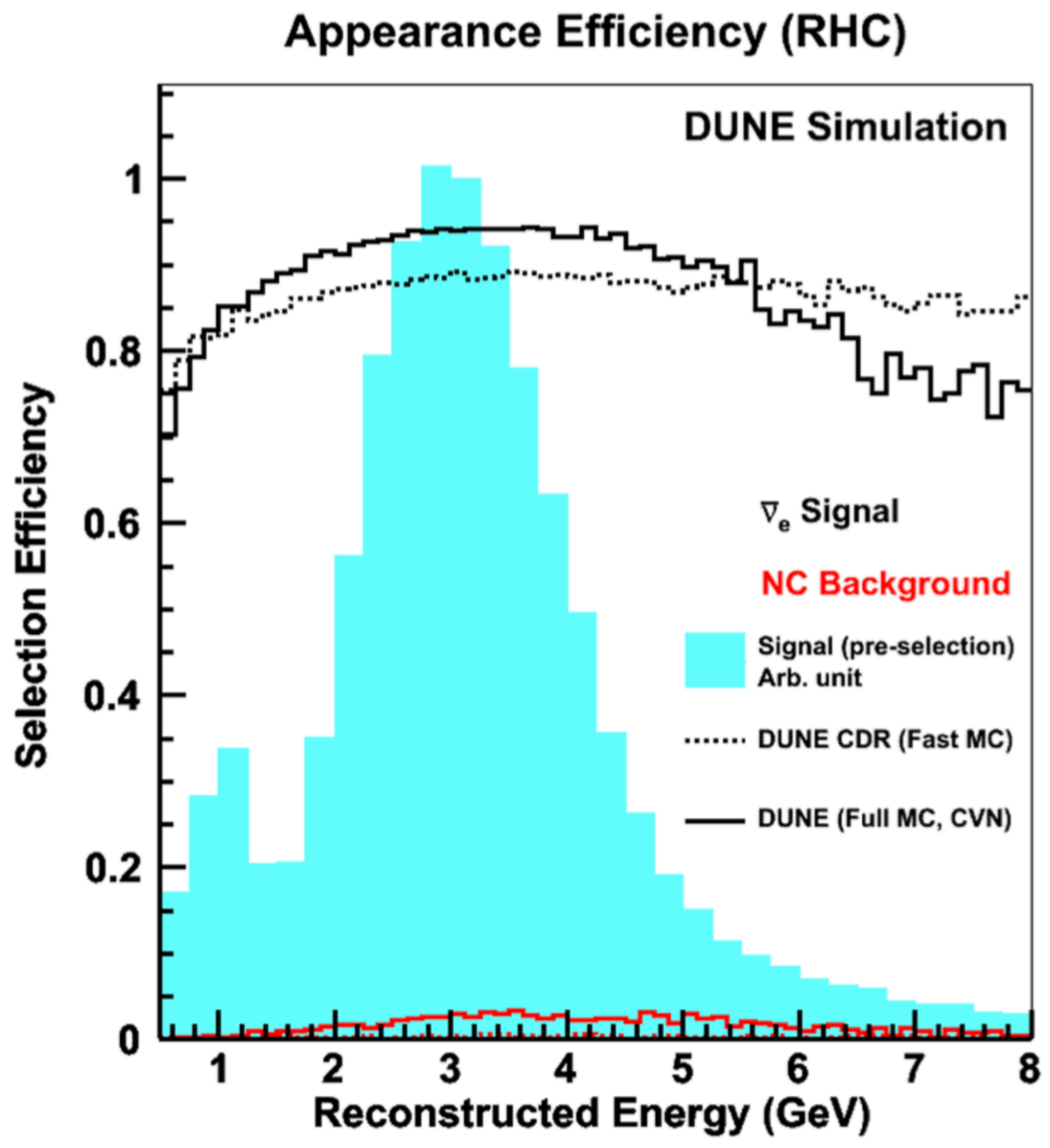
DUNE CVN

- We obtain highly efficiency analyses from the CVN event selection

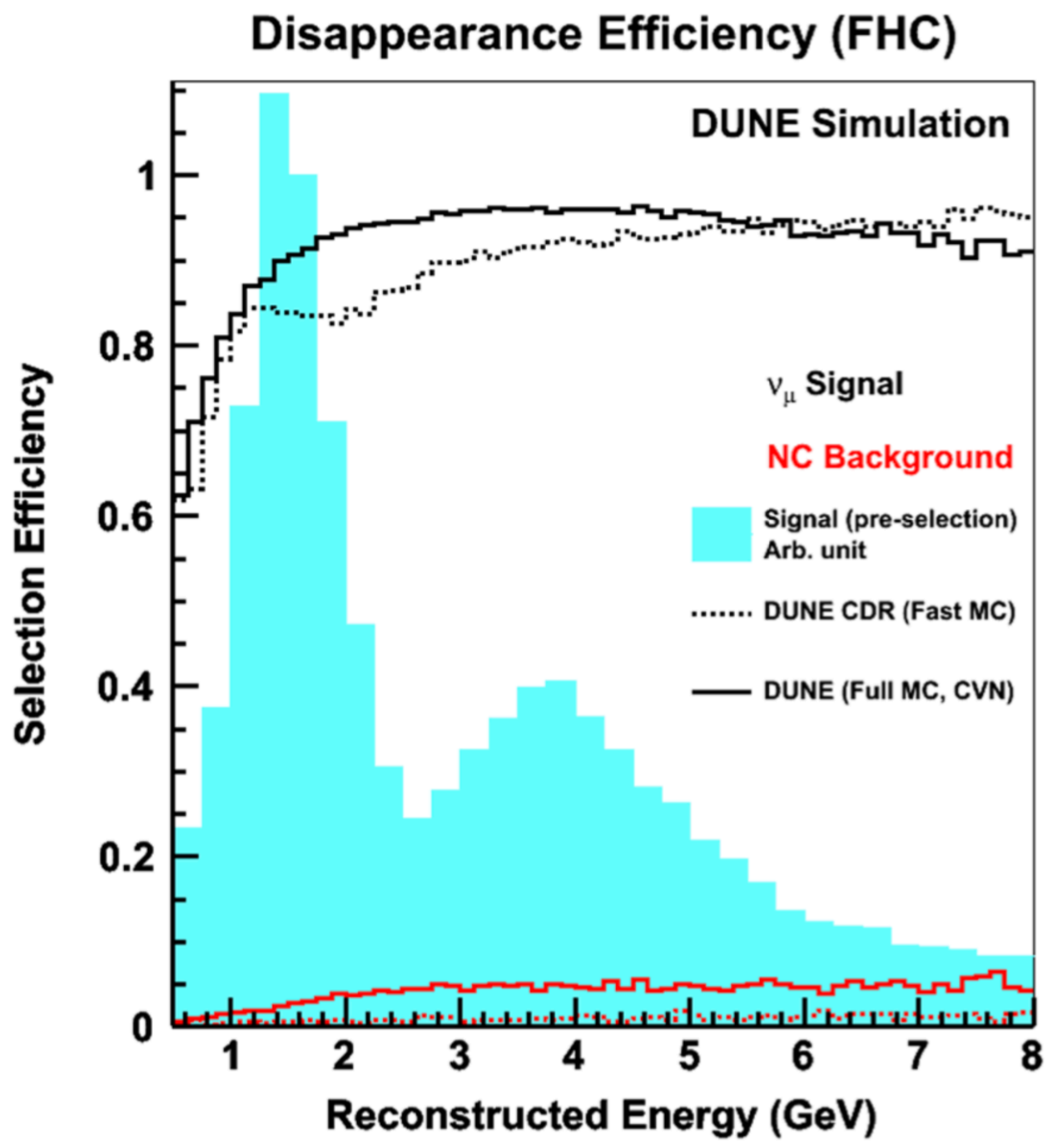
Neutrino mode



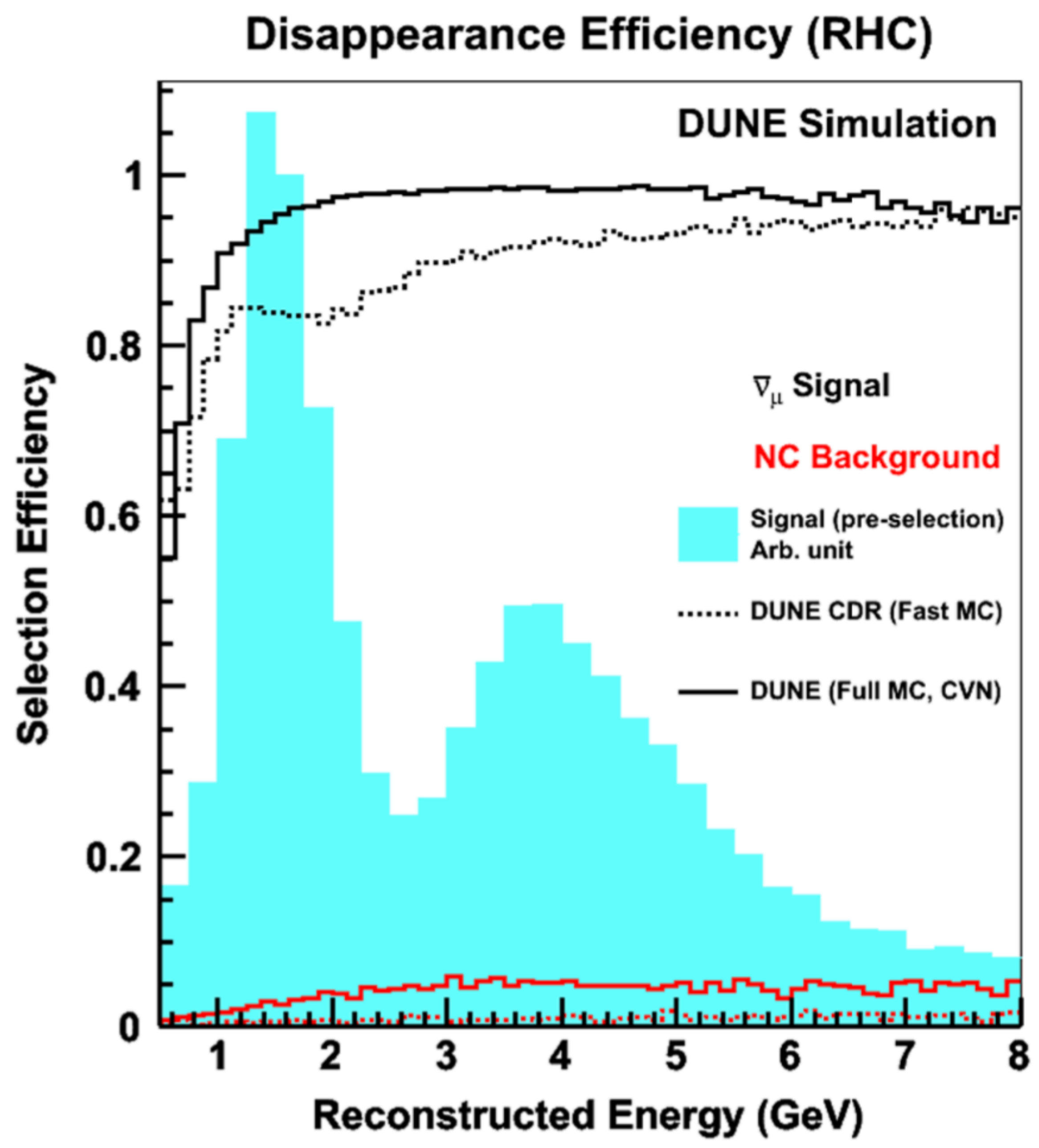
Antineutrino mode



Neutrino mode



Antineutrino mode



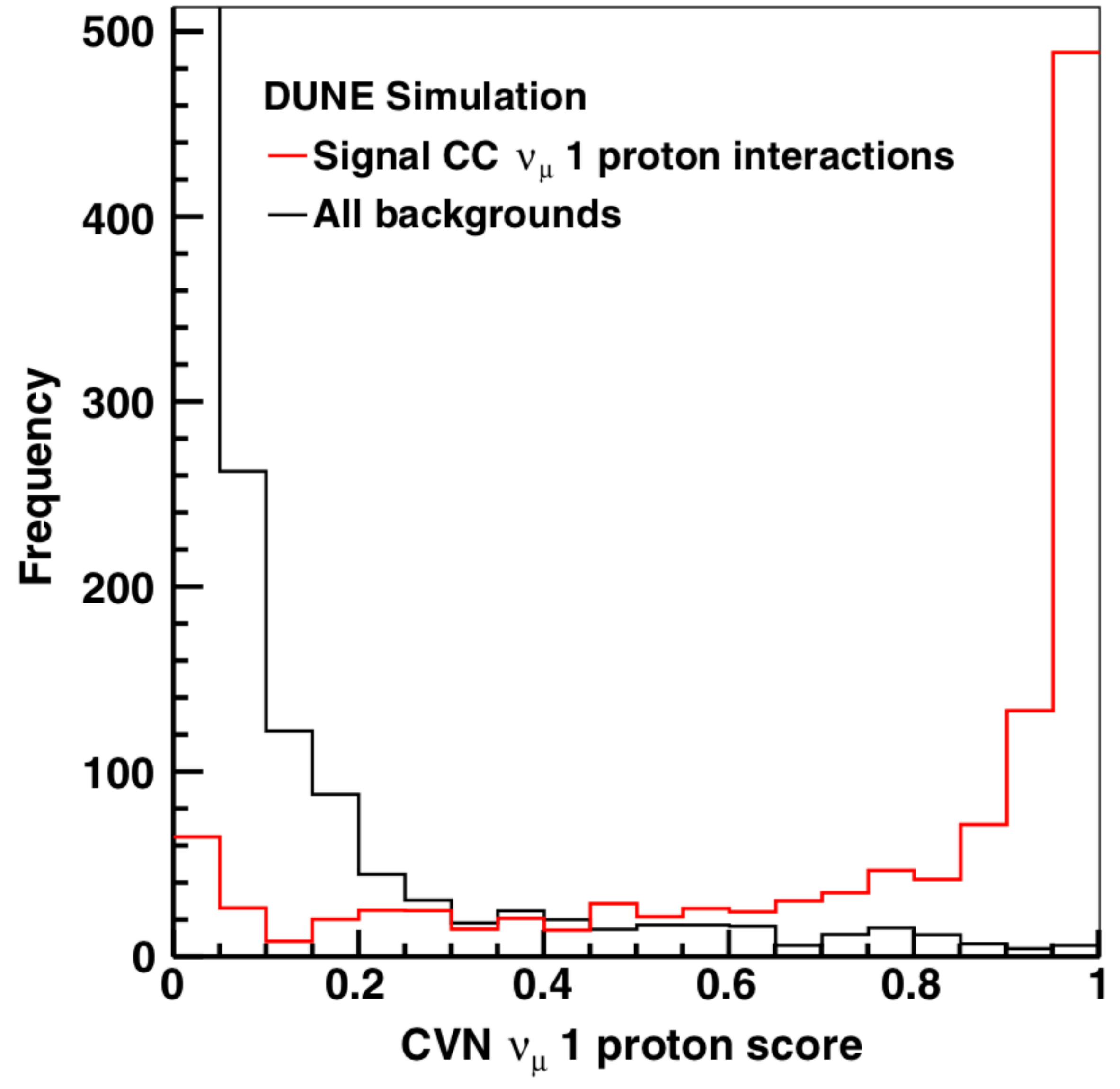
Efficiency for selecting CC ν_e interactions

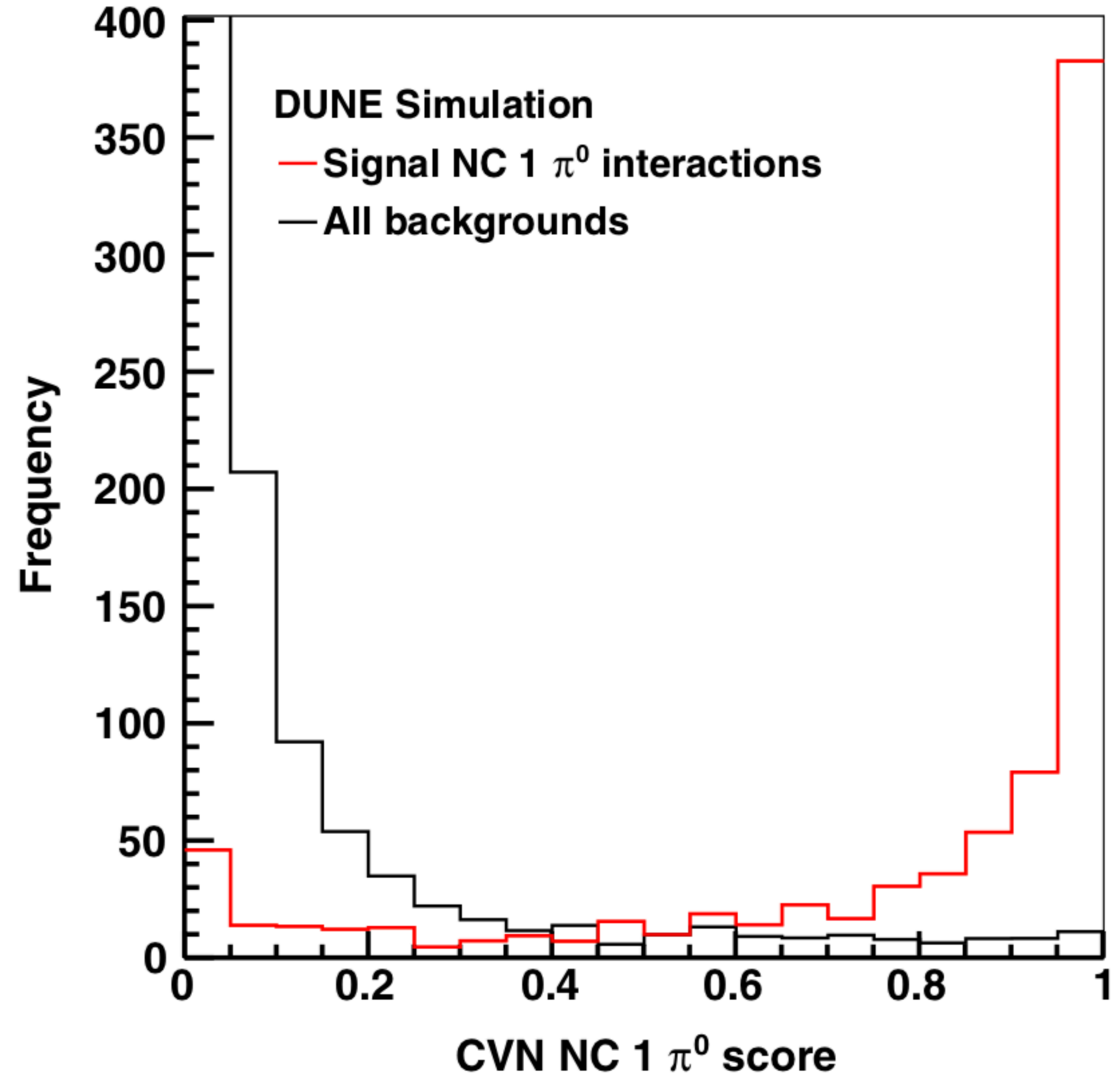
Efficiency for selecting CC ν_μ interactions

[1] DUNE Collaboration, *Neutrino interaction classification with a convolutional neural network in the DUNE far detector,* Phys.Rev.D 102 (2020) 9, 092003.

DUNE CVN - Particle counting

- We tested some of the particle counting outputs
  - Proof of principle of using the CVN for exclusive final state selections
- Multiply together different scores:
  - CC ν_μ, 1p, $0\pi^\pm$, $0\pi^0$
  - NC, 0p, $0\pi^\pm$, $1\pi^0$





- Clearly these would need to be strongly validated before use on data
  - Much more likely to be biased by the choice of event generator

[1] DUNE Collaboration, *Neutrino interaction classification with a convolutional neural network in the DUNE far detector*, Phys.Rev.D 102 (2020) 9, 092003.

Transfer Learning

- Transfer learning makes use of previously trained networks
  - Allows you to fine tune a pre-trained network for your task
  - Can be useful if you don't have much data
  - The idea dates back to the early days of perceptrons[1]

- I will discuss a recent study we performed on using transfer learning in neutrino event classification

Eur. Phys. J. C (2022) 82:1099
https://doi.org/10.1140/epjc/s10052-022-11066-6

THE EUROPEAN PHYSICAL JOURNAL C

Check for updates

Regular Article - Experimental Physics

Application of transfer learning to neutrino interaction classification

Andrew Chappell[2,a], **Leigh H. Whitehead**[1,b]

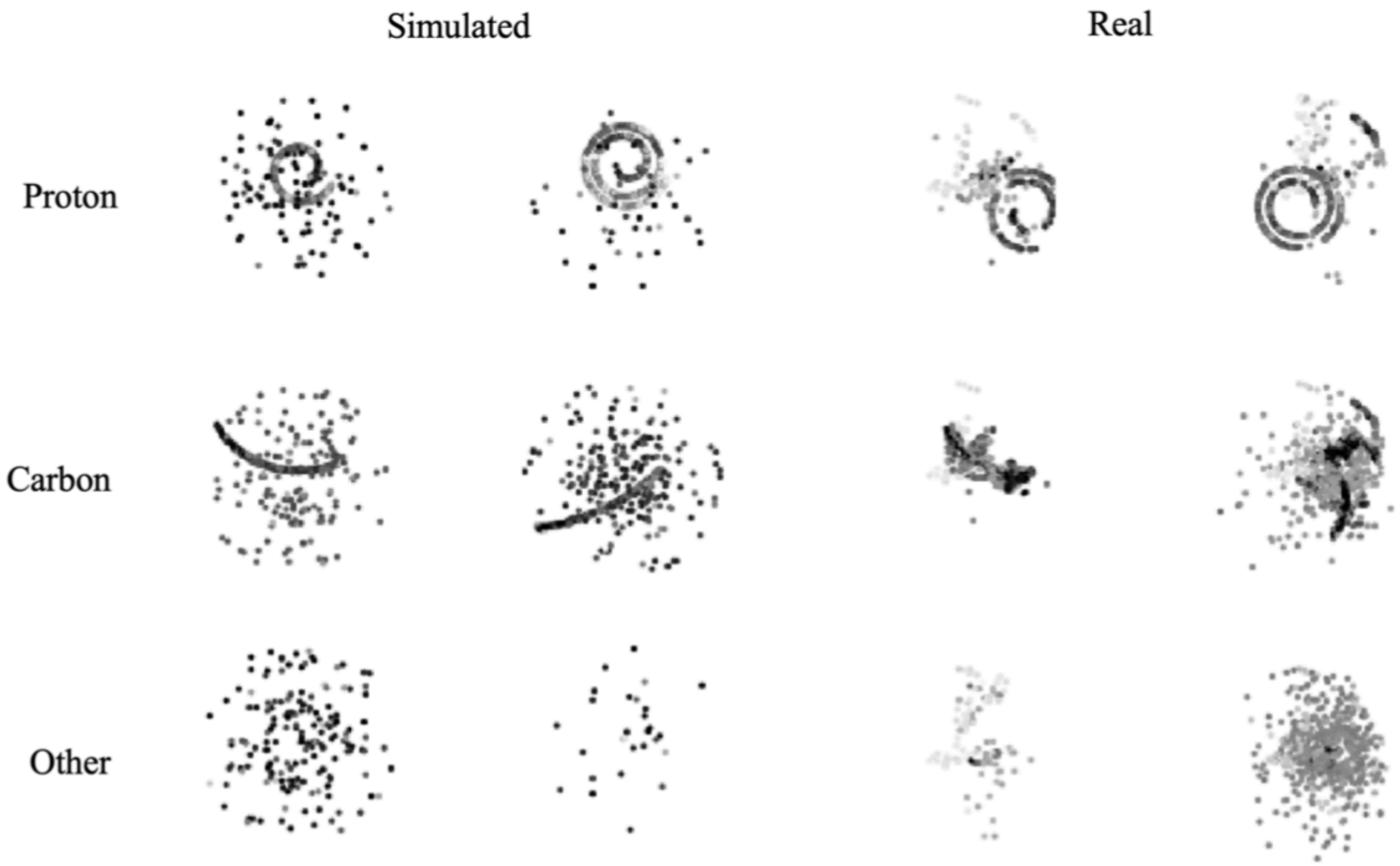
[1] Department of Physics, University of Cambridge, Cambridge CB3 0HE, UK
[2] Department of Physics, University of Warwick, Coventry CV4 7AL, UK

https://link.springer.com/article/10.1140/epjc/s10052-022-11066-6

[1] S. Bozinovski, A. Fulgosi, *The influence of pattern similarity and transfer learning upon the training of a base perceptron* b2. In: Proceedings of Symposium Informatica, Bled, Slovenia (1976) p. 3–1215.

Transfer Learning in Physics

- I was only able to find once example of transfer learning in a related field when we started this work
- The AT-TPC[1] was a nuclear physics experiment
- Used transfer learning due to a small simulation dataset
- Also used some hand-labelled data due to poor simulation quality



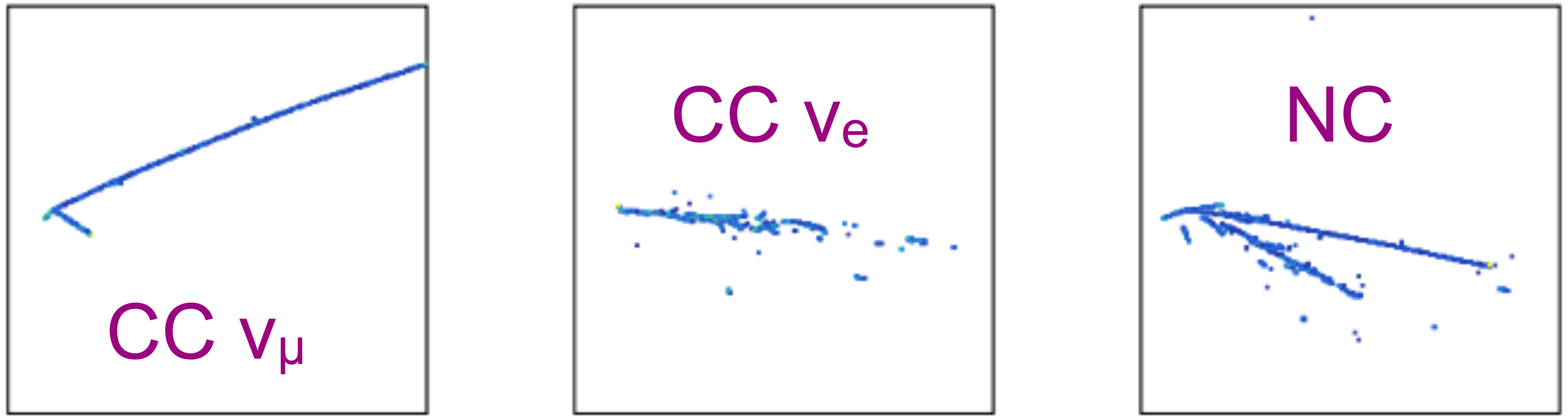
[1] M. P. Kuchera, et al., Machine Learning Methods for Track Classification in the AT-TPC, NIM A 940 (2019) 156-167, 1810.10350

Transfer Learning in LArTPCs

- Can we use transfer learning to reduce the number of training examples?
  - Simulations are time consuming and GPUs need a lot of power
- Conveniently, LArTPC detectors, such as DUNE, have three readout planes
  - We get three images of a given interaction
  - Photographic images have depth three (red, green and blue channels)
- Can we use a network trained on photographs for our event classification?
  - There are plenty of networks trained on photograph-based challenges
  - Use these networks as a starting point and fine tune the weights

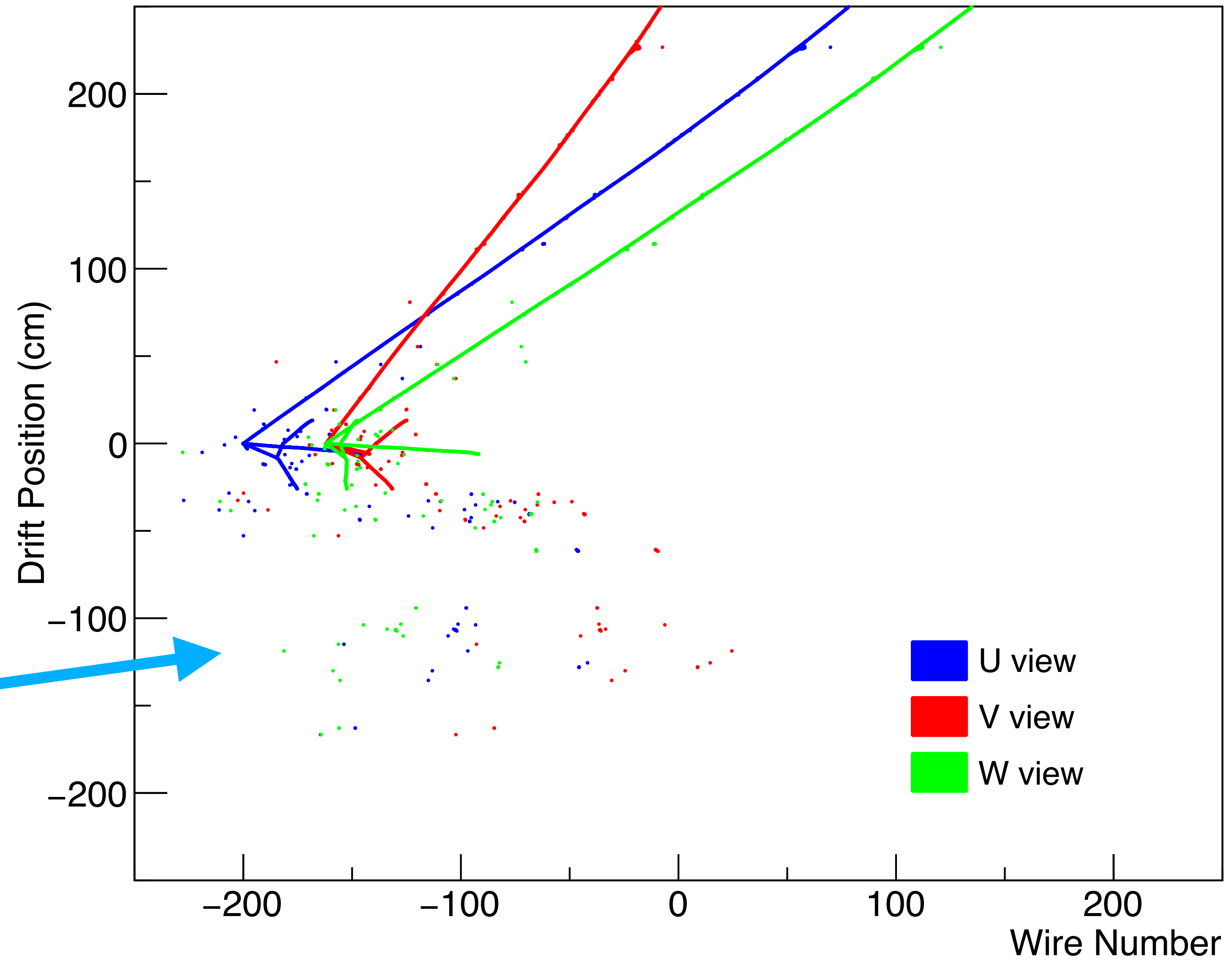
TL: Event Sample

- GENIE neutrino events:
  - CC ν_μ, CC ν_e and NC
  - 50,000 of each type



- Events passed through simple LArTPC simulation
  - Outputs three images of each event
  - Three projections of the (y,z) plane

CC ν_μ event with the three views overlaid as RGB channels



TL: Architecture and Training

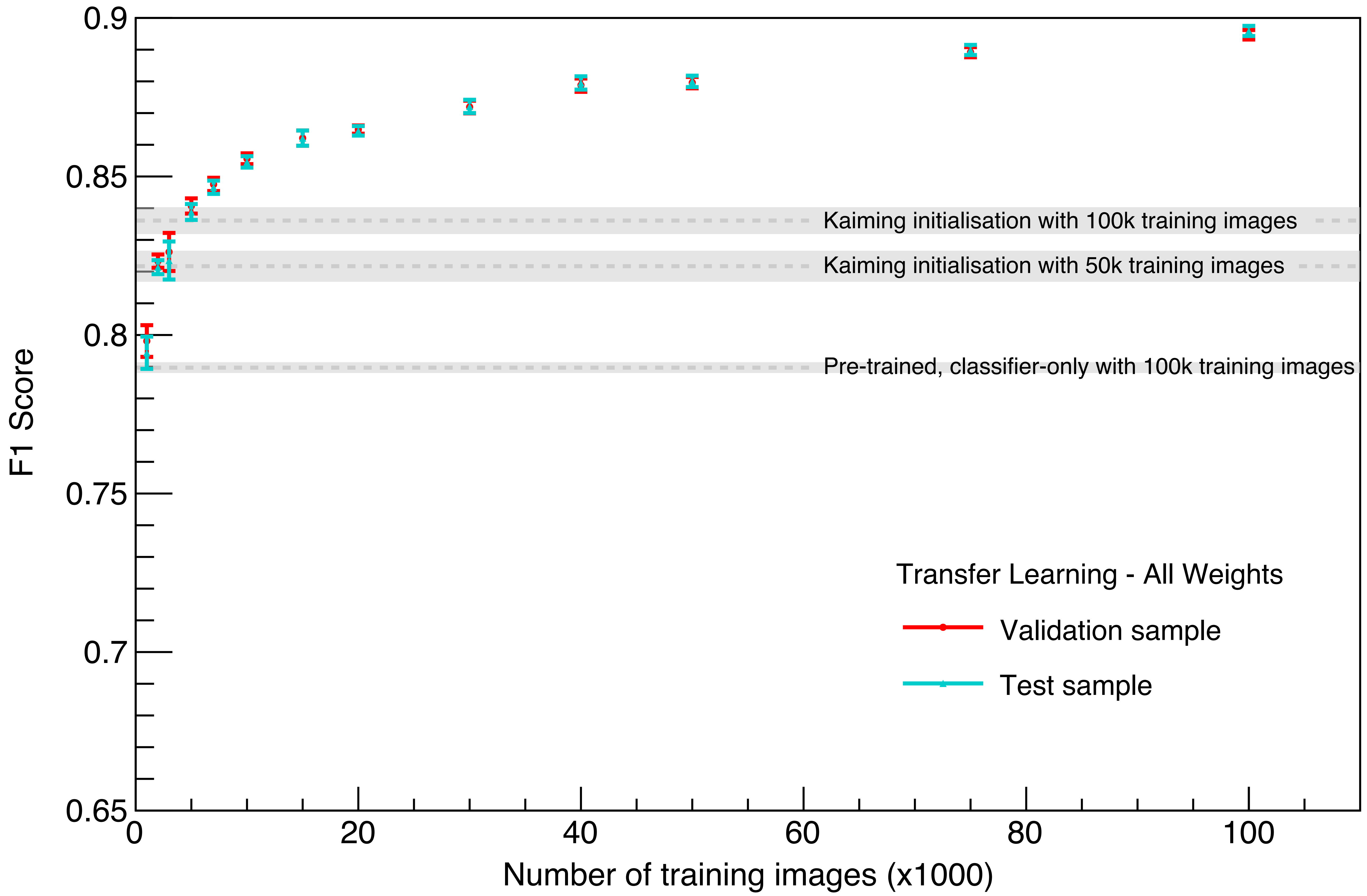
- We chose to use the Pytorch implementation of ResNet18
  - Small depth was chosen since this study involved training over 1000 networks
- The pre-trained version of ResNet18 was trained on ImageNet
  - We had to change the final layer from 1000 to 3 classes: CC ν_{μ}, CC ν_{e} and NC
- Trained a series of networks with:
  - Kaiming (He) randomly initialised weights
  - Weights from the pre-trained ImageNet network
  - Various numbers of training events from 1,000 to 100,000
  - Trained each network 25 times to give an estimate of the uncertainty

Results: TF vs random initialisation

- Compared the F1 score from the transfer-learned networks fine-tuned with 1k to 100k images against the Kaiming-initialised network with 50k and 100k events

- Transfer-learned network out-performs the Kaiming-initialised network with 100k training images
  - For 7k training images and above

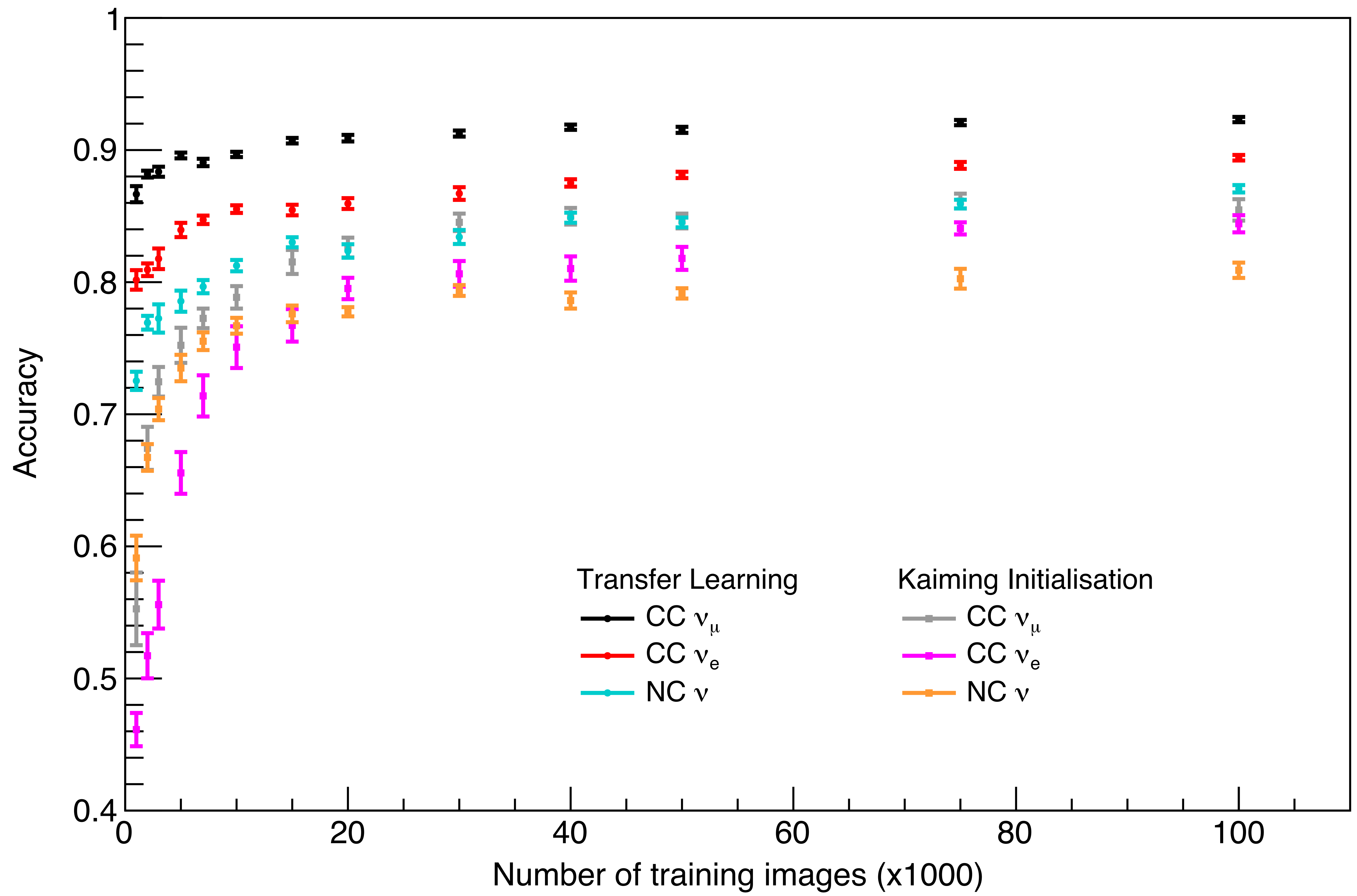
- Event fine-tuning just the final layer works surprising well
  - F1 score = 0.79



Results: TF vs random initialisation

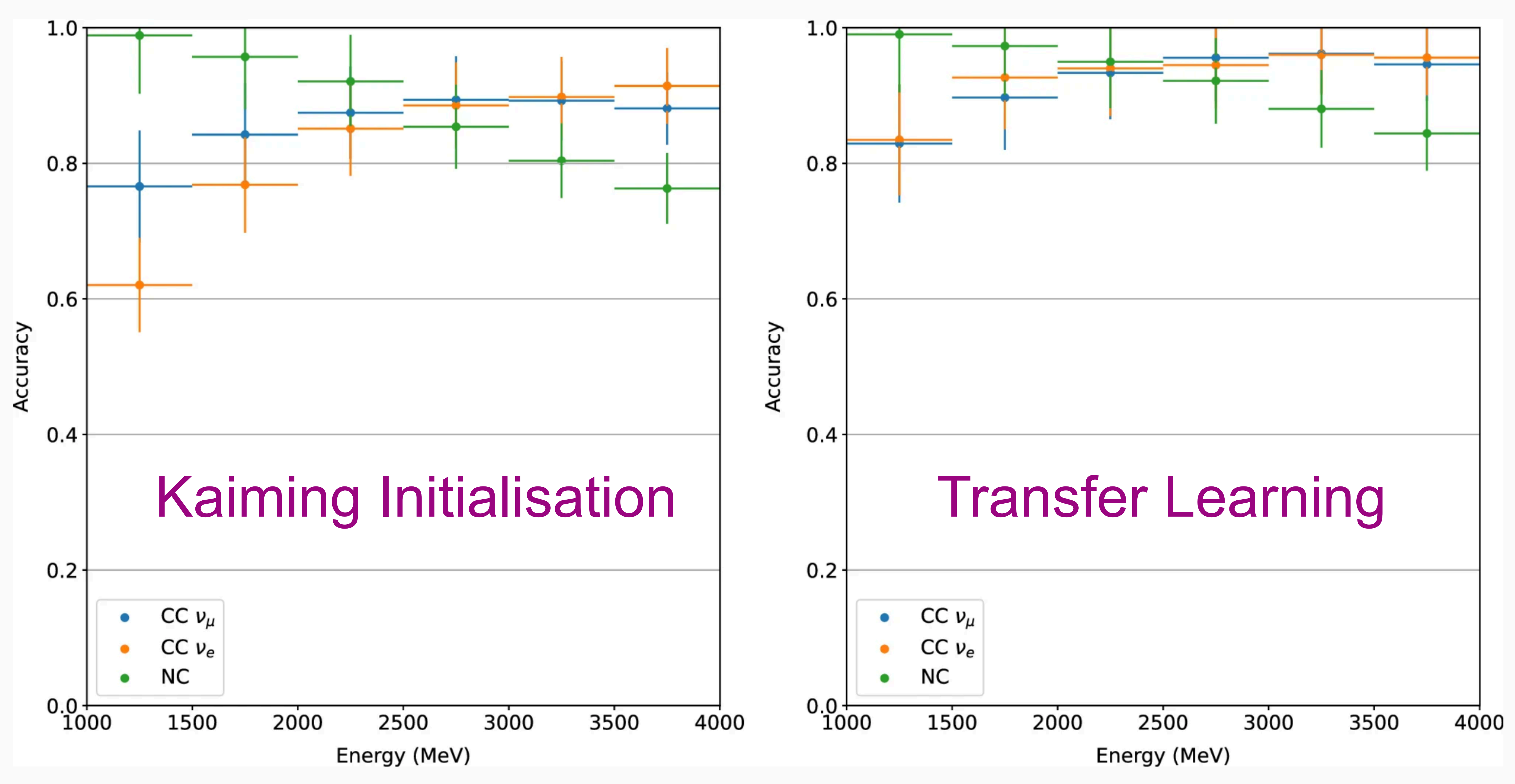
- Compared the F1 score from the transfer-learned networks fine-tuned with 1k to 100k images against the Kaiming-initialised network with 50k and 100k events

- Improvement is seen in all three classes (not just overall)



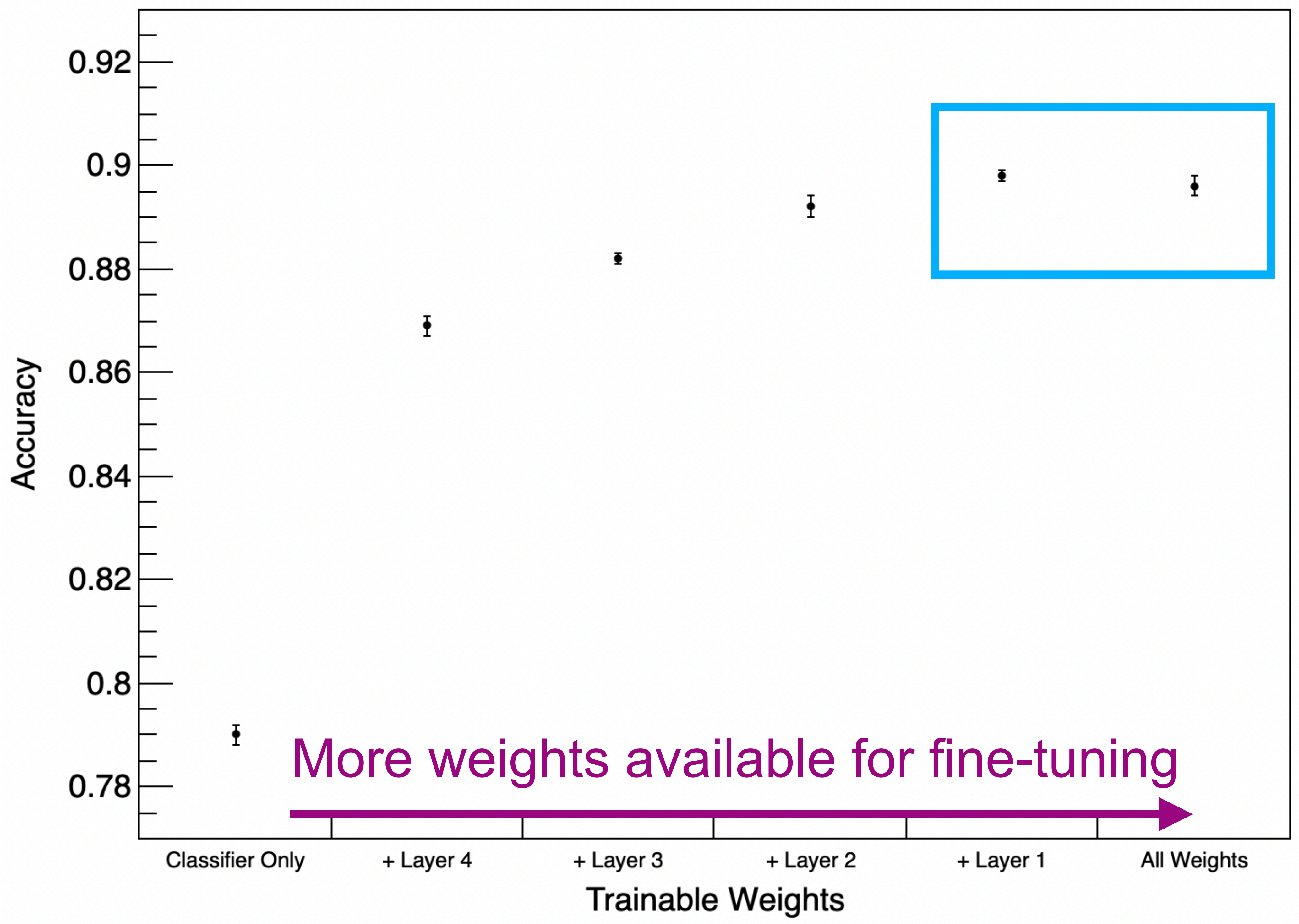
Transfer Learning in LArTPCs

- We also looked for potential biases between classes and as a function of energy
  - See reduced bias in both cases using transfer learning
  - Plots show examples from training with 100k events



Transfer Learning in LArTPCs

- Also looked at the effect of freezing different layer weights
  - Layers 1 to 4 here correspond to the ResNet blocks
  - As a minimum we have to train the classifier (dense layer)
  - The difference between Layer 1 and All Weights is the first convolutional layer
    - No difference in performance is seen when the first layer weights can be fine-tuned
    - The ImageNet-trained first convolutional layer extracts all the information needed to classify our neutrino events



Conclusions

- Use of CNNs for neutrino event classification is now well-established
- Transfer learning looks to be a promising approach
  - Can perform very well with small training samples
  - Reduces number of required training examples
  - Reduced bias for smaller training samples
  - Training process seems more stable
  - The first convolutional layers from networks trained on photographs can extract all of the required low-level features

Thank you… any questions?