



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

Received: 8 August 2022 / Accepted: 23 November 2022
© The Author(s) 2022

Abstract Training deep neural networks using simulations typically requires very large numbers of simulated events. This can be a large computational burden and a limitation in the performance of the deep learning algorithm when insufficient numbers of events can be produced. We investigate the use of transfer learning, where a set of simulated images are used to fine tune a model trained on generic image recognition tasks, to the specific use case of neutrino interaction classification in a liquid argon time projection chamber. A ResNet18, pre-trained on photographic images, was fine-tuned using simulated neutrino images and when trained with one hundred thousand training events reached an F1 score of 0.896 ± 0.002 compared to 0.836 ± 0.004 from a randomly-initialised network trained with the same training sample. The transfer-learned networks also demonstrate lower bias as a function of energy and more balanced performance across different interaction types.

1 Introduction

The usage of Deep Learning has increased rapidly in neutrino physics over the last 5–10 years [1,2]. The data from many neutrino experiments can be easily and naturally represented in an image format, hence Convolutional Neural Networks (CNNs) are a very popular choice of deep learning algorithm in the field. CNN models contain millions of parameters that must be trained, which is typically done using large numbers of simulated neutrino interactions. For example, the CNN used to perform the neutrino event classification in the Deep

Underground Neutrino Experiment (DUNE) was trained on over three million simulated events [3,4].

However, detector simulations for large detectors are very time consuming and resource intensive, so other methods are being explored to be able to train powerful and accurate deep learning algorithms without a very large computational burden. Potential solutions to this problem fall in to three categories: methods to make faster simulations, methods to improve computational performance of the networks and methods to reduce the number of simulated events required. However, the use of GPUs in deep learning can carry its own computational burden, and this resource intensity is becoming of increasing importance in the light of high energy costs and increased focus on the carbon footprint of research activities [5] and we must therefore ensure we use such resources as effectively and efficiently as possible. Methods to make faster simulations often use a generative model, typically a Generative Adversarial Network, to approximate the simulation to produce events much more quickly (see, for example, Chapter 6 of Ref. [6] for a review). To improve computational performance, alternative network architectures using sparse representations of the images have been deployed (see, for example, Ref. [7]). For reduction of event requirements *transfer learning* can be used as an approach to use a much smaller number of simulated events to fine tune an existing, pre-trained model, with these models often trained on photographic images.

Transfer learning was first proposed in 1976 by Bozinovski and Fulgosi [8,9] for the training of perceptrons. More recently it has been applied to deep learning [10], including in fields similar to neutrino physics: an example from the AT-TPC nuclear physics experiment showed that a fairly small number (thousands) of training examples gave good performance when used to fine tune a generically pre-trained model [11].

Andrew Chappell and Leigh H. Whitehead have contributed equally to this work.

^a e-mail: andrew.chappell@warwick.ac.uk (corresponding author)

^b e-mail: leigh.howard.whitehead@cern.ch (corresponding author)

As in the AT-TPC experiment, liquid argon time projection chamber (LArTPC) event displays bare little resemblance to the photographic images used to train existing models, and therefore the goal of this article is to assess the effectiveness of using transfer learning for the classification of interactions in a DUNE-like LArTPC detector and determine the most appropriate approaches to fine tuning. The details of the event simulation and image production are given in Sect. 2, Sect. 3 presents a case-study with the aim of classifying three general types of neutrino interactions, Sect. 4 presents the results of the study, and Sect. 5 provides a discussion and closing remarks.

2 Simulated event samples

Neutrino interactions were generated using GENIE $v3_00_06$ [12] and a uniform flux distribution between $1 \text{ GeV}/c^2$ and $4 \text{ GeV}/c^2$. The flux distribution was chosen to give a rough approximation of the DUNE flux in the main oscillation region of the spectrum [13]. Three balanced samples of interaction were produced: charged-current muon neutrino (CC ν_μ), CC electron neutrino (CC ν_e), and neutral current (NC) interactions. The important outputs from GENIE in this case are the kinematics of the incoming neutrino and the argon target, and the kinematics of all of the final-state particles produced in the interaction.

The final-state particles are tracked through a simple LArTPC detector using Geant4 $v4_10_6$ [14]. The detector geometry is defined as a cuboid filled with liquid argon and the dimensions in the (x, y, z) directions are $5 \text{ m} \times 5 \text{ m} \times 5 \text{ m}$, where z defines the beam direction, y is vertical and x is the drift direction. The simulation produces three-dimensional energy deposits within the detector volume that are projected into three two-dimensional views of the yz plane, similar to the three wire readout planes in the planned DUNE detectors [15]. These three views are referred to as u , v and w and are aligned at 35.9° , -35.9° and 0° to the vertical, respectively.

The output of the simulation is formed by three two-dimensional images showing u , v , and w on the horizontal axis, and the drift coordinate x on the vertical axis. The pixel intensity is given by the amount of energy deposited. Examples of a CC ν_μ , a CC ν_e and an NC interaction are shown in the left, centre and right panels of Fig. 1, respectively. Each event is shown in the (w, x) view. The CC ν_μ event shows the characteristic long muon track, the CC ν_e event shows the typical electron shower emanating from the interaction vertex, and the example NC event shows that NC events can sometimes include components similar to the CC ν_μ and CC ν_e interactions. The chosen pre-trained network requires 224×224 pixel input images, so the images are produced such that each pixel represents a $1 \times 1 \text{ cm}^2$ region, cropped

and centred on the region surrounding the interaction. This represents an approximately twofold decrease in resolution compared to the $\sim 5 \text{ mm}$ granularity of the readout planes in order to contain larger interactions within the images. These pre-trained networks are used to classify photographic images, hence the three images of each event are stacked together to produce a depth three image that is analogous to a colour image with red, green and blue colour channels.

The total number of images available was 140,000, of which 20,000 events were used as a validation set and another 20,000 images as a test set to produce the final results.

3 Event classification case study overview

The aim of this study is to investigate the use of transfer learning to train a CNN for the task of neutrino event classification. In the simplest case, long-baseline neutrino oscillation experiments need to be able to accurately and efficiently identify CC ν_μ , CC ν_e , and NC interactions. Each neutrino interaction will therefore be classified as one of the three true categories: CC ν_μ , CC ν_e or NC.

The PyTorch [16] framework was used because of the wide range of pre-trained architectures available. The architecture chosen was the ResNet18 [17], since ResNets are a popular choice in the neutrino physics field and the relatively shallow depth eases the computational burden for training the hundreds of networks required by this study.

3.1 Network architecture

We consider the architecture as two sub-networks: the *feature extractor* network and the *classifier* network. The feature extractor network consists of the many convolutional layers that extract features from the input images, and the classifier network, that provides the specific outputs for the task being performed. The classifier is specific to each use case, so must be appended to the predefined ResNet18 feature extractor network. The choice for the classifier was a single three node dense layer taking the $(n_b, 512)$ output from the ResNet18 and returning the final three classification scores, where n_b is the number of images per batch. The final architecture is shown in Fig. 2, where only the convolutional and pooling layers of the ResNet18 have been shown for clarity. The naming convention from the ResNet architecture is used here, which breaks up the network into six blocks of layers, with the middle four blocks named Layer 1 to Layer 4. The total number of trainable weights in the network is 11,178,051, divided between the layers as follows: 9536 (Layer 0); 147,968 (Layer 1); 525,568 (Layer 2); 2,099,712 (Layer 3); 8,393,728 (Layer 4); 1539 (classifier).

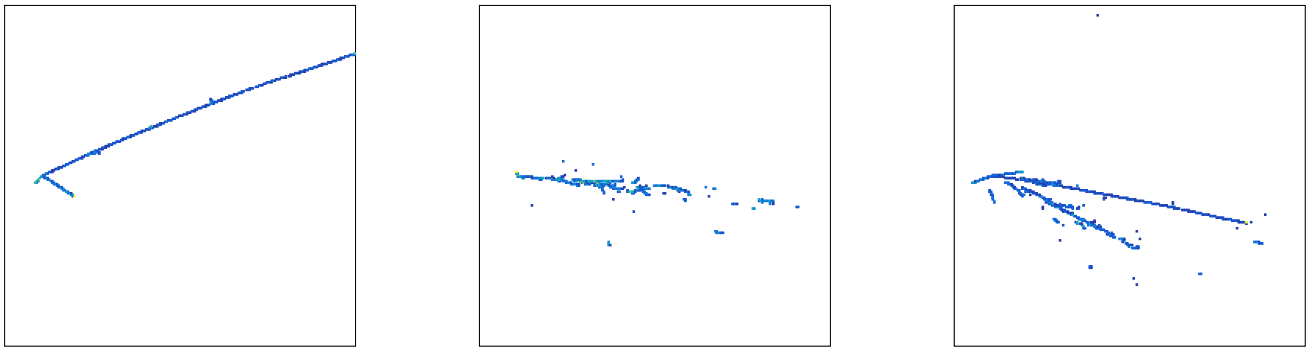


Fig. 1 Three simulated neutrino interactions shown in the (w, x) view. Examples of a CC ν_μ , a CC ν_e and an NC interaction are shown in the left, centre and right panels, respectively. Each image is 224×224 pixels,

each pixel corresponds to an area of $1 \times 1 \text{ cm}^2$, and the colour represents the deposited energy (blue is lowest, yellow is highest)

3.2 Training details

Stochastic gradient descent (SGD) was chosen as the optimiser. In all cases, the starting learning rate was set to 0.001, and it was reduced by a factor of 10 each time the validation loss did not improve for three epochs. The network training stopped automatically when the validation loss did not reduce for six epochs. A batch size of 32 images was chosen as an optimisation between classification performance, training time and memory usage. All networks were trained using a NVIDIA Tesla V100.

3.3 Performance metrics

The F1 score [18] can be written in terms of true positives ($T_{p,i}$), false positives ($F_{p,i}$) and false negatives ($F_{n,i}$), where the suffix i indicates the target class under consideration. For a multi-class classification with n classes, such as in the analysis presented here with $n = 3$, the overall F1 score can be calculated from the individual scores for each class in a number of ways. Here the *macro* averaging scheme is used, such that the metric is computed on a per-class basis and then the F1 score is taken to be the unweighted mean of the per-class scores:

$$F_1 = \frac{1}{3} \sum_{i=1}^3 \frac{T_{p,i}}{T_{p,i} + 0.5(F_{p,i} + F_{n,i})} \quad (1)$$

Equation (1) shows that the allowed values of the F1 score are between zero and one, where one is the perfect score when there are no false positives or false negatives ($F_{p,i} = F_{n,i} = 0$). It also considers false positives and false negatives as equally bad in terms of calculating the score.

For each study presented, an ensemble of 25 independently trained versions of the network was produced, with the mean and error on the mean of the different metrics forming the reported results. The use of ensembles accounts for

random fluctuations in the initialisation and training of the networks that arise from the fact that these are stochastic processes. It is important to note that when using a fixed random seed the results are deterministic and reproducible on a given system.

4 Results

4.1 Randomly initialised ResNet18

ResNet18s with random weight initialisation form the baseline for this study against which the transfer learning results will be compared in Sect. 4.4. The Kaiming (also known as He) [19] initialisation scheme was developed specifically for CNNs using non-linear activation functions such as ReLU. The specific version of the Kaiming initialisation scheme used in this work was the normal distribution form. The standard deviation of the distribution depends on the number of weights in the layer (equivalent to the size of the output from the previous layer), n_w : $\sigma = \sqrt{2/n_w}$.

An ensemble of 25 networks were trained with differing numbers of training events used: 1000 (1k); 2k; 3k; 5k; 7k; 10k; 15k; 20k; 30k; 40k; 50k; 75k; and 100k. Table 1 shows the F1 scores from the testing and validation samples from the networks trained with the above number of interactions. As expected, the performance increases significantly as the number of training events rises. The uncertainty on the F1 score is seen to reduce as a function of the number of training images, which is expected as the training should become more stable with more training examples. Using the full training dataset of 100k events, an F1 score of 0.836 ± 0.004 was measured.

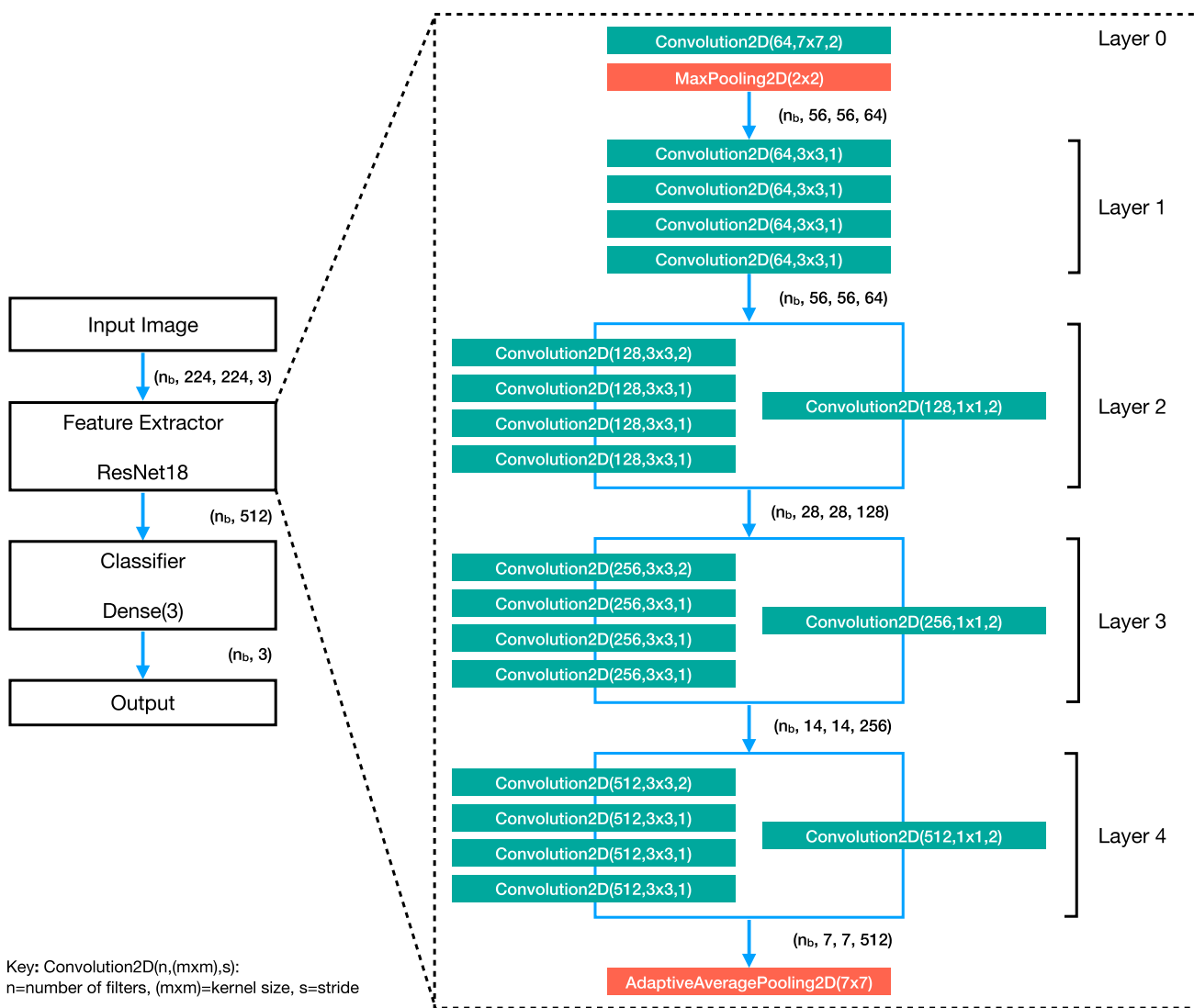


Fig. 2 An overview of the architecture, where 224×224 pixel images of depth three are input to the ResNet18 feature extractor that outputs a 512 element tensor. The classifier takes the 512 element tensor and outputs three classification scores, one for each class. The number of images per batch, n_b , is shown in the tensor dimensions. For clarity, only the convolutional and pooling layers are shown for the feature extractor.

The convolutional layers with a stride of two perform a downsampling of a factor of two in the image size, and the residual connections are those with the single 1×1 convolutional layers. Following the ResNet convention, the convolutional layers are combined into groups called Layers 1 to 4

4.2 Transfer learning with ResNet18

The pre-trained ResNet18 that forms the basis of this study was trained on the ImageNet [20] data sample, meaning that it was trained on photographic images with the goal of classifying them into one of one thousand categories. The classifier network was modified to provide the three required outputs for this use case, as described in Sect. 3.1.

Samples of neutrino interaction images were then used to fine tune the weights of the pre-trained networks. Different networks have been trained with different numbers of training images, ranging from one thousand to one hundred thousand

(with approximately equal fractions from each of the three true classes). The performance has been studied as a function of the number of ResNet18 Layers (as defined in Fig. 2) with weights that are allowed to be fine-tuned, where the weights of the ResNet18 layers were progressively frozen:

- *All Weights*: No weights frozen, total of 11,178,051 trainable weights.
- *Freeze(1)*: Layer 0 weights frozen, total of 11,168,515 trainable weights remain.
- *Freeze(2)*: Layer 0 and 1 weights frozen, total of 11,020,547 trainable weights remain.

Table 1 The F1 scores for the testing sample for the Kaiming initialised networks. The values and uncertainties given come from the mean and error on the mean of 25 different training attempts

Model	Initialisation	Training images	Testing F1 score
ResNet18	Kaiming	1k	0.528 ± 0.014
		2k	0.621 ± 0.010
		3k	0.661 ± 0.010
		5k	0.720 ± 0.010
		7k	0.747 ± 0.009
		10k	0.770 ± 0.008
		15k	0.787 ± 0.008
		20k	0.800 ± 0.004
		30k	0.816 ± 0.005
		40k	0.816 ± 0.004
		50k	0.822 ± 0.005
		75k	0.833 ± 0.004
100k	0.836 ± 0.004		

- *Freeze(3)*: Layer 0, 1 and 2 weights frozen, total of 10,494,979 trainable weights remain.
- *Freeze(4)*: Layer 0, 1, 2 and 3 weights frozen, total of 8,395,267 trainable weights remain.
- *Classifier Only*: All ResNet18 weights frozen, total of 1,539 trainable weights remain.

When using a pre-trained feature extractor, it is clear that the classifier weights are the only ones that *must* be trained in order to get performance better than random guessing. Beyond this, training Layer 4 will likely give the biggest step in performance because it contains approximately 75% of the network weights. Generally, it is expected that the performance will improve as the number of trainable parameters increases. Furthermore, of interest to this use case is the evolution of the features that can be extracted at each layer of the network. Zeiler and Fergus [21] showed that early CNN layers comprise low-level geometric features (edges, corners, etc), with deeper layers becoming increasingly more class-specific. The expectation therefore, is that early layers may retain a high degree of relevance when applied to this use case, while deeper layers will increasingly contain many feature extractors of little relevance to our use case. One might therefore expect that fine tuning can be limited to deeper layers, which an assessment of performance by layer will also determine.

4.3 Comparison of transfer learning cases

The top section of Table 2 shows the comparison between the networks that had different numbers of weights free for fine-tuning when trained using the full sample of 100,000 interactions. As expected, the performance is best when allowing more weights to be fine-tuned since it gives the network more degrees of freedom to perform the classification. Within sta-

tistical uncertainties, the results from the All Weights and Freeze(1) categories are the same, which is to be expected since there are very few parameters in the first convolutional layer of the ResNet18. The best F1 score is hence reported as 0.896 ± 0.002 for fine tuning all of the network parameters. The fact that fine tuning the initial convolutional layer has very little effect on the CNN performance suggests that even though it was trained on photographic images, it is extracting generic features that are applicable to the LArTPC images. It is notable that training only the classifier weights still obtains a F1 score of 0.790 ± 0.002 , which, when compared to the results in Table 1, outperforms training the Kaiming-initialised network with fewer than 20k images. Furthermore, the addition of only the Layer 4 weights is sufficient to yield an F1 score of 0.869 ± 0.002 , out-performing the Kaiming-initialised network with the full 100k images.

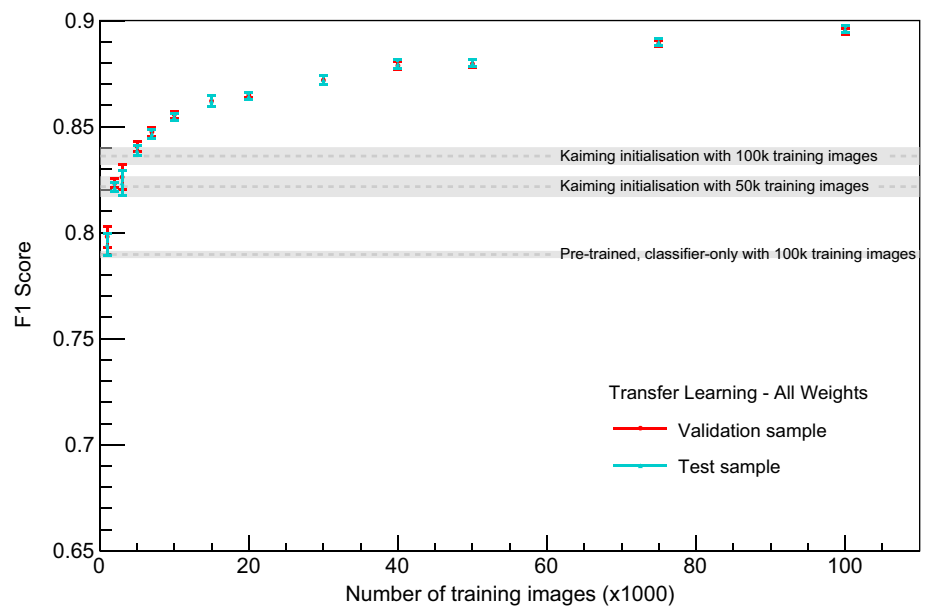
4.4 Comparison of transfer learning and random initialisation

The F1 score measured for the transfer learning all weights case as a function of the number of training images is shown in the bottom section of Table 2. Even when trained with 1k events it outperforms the classifier only case with 100k training examples. The results are shown graphically and compared to the randomly-initialised Kaiming networks for 50k and 100k training samples in Fig. 3. It shows that, in the case of fine tuning all of the network weights (cyan points for testing sample, red points for validation sample), the performance exceeds the Kaiming-initialised ResNet18 trained on 100k (50k) images using only 7k (5k) images. This is a powerful demonstration of the use of transfer learning even when a reasonably large training sample of 100k events is available to train a randomly-initialised CNN. Using all 100k events in the transfer learning case improves the F1 score

Table 2 The testing F1 scores for various combinations of the transfer learning cases and number of training images. The reported values show the mean value of the 25 trained networks and the associated error on the mean

Model	Initialisation	Weights scheme	Training images	Testing F1 score
ResNet18	Pre-trained	All Weights	100k	0.896 ± 0.002
		Freeze(1)	100k	0.898 ± 0.001
		Freeze(2)	100k	0.892 ± 0.002
		Freeze(3)	100k	0.882 ± 0.001
		Freeze(4)	100k	0.869 ± 0.002
		Classifier Only	100k	0.790 ± 0.002
ResNet18	Pre-trained	All Weights	1k	0.794 ± 0.005
			2k	0.821 ± 0.002
			3k	0.824 ± 0.006
			5k	0.839 ± 0.003
			7k	0.847 ± 0.002
			10k	0.855 ± 0.002
			15k	0.862 ± 0.002
			20k	0.864 ± 0.002
			30k	0.872 ± 0.002
			40k	0.880 ± 0.002
50k	0.880 ± 0.002			
75k	0.890 ± 0.002			
100k	0.896 ± 0.002			

Fig. 3 The test sample (cyan) and validation sample (red) F1 scores as a function of the number of images used to train the all-weights transfer learning networks. The three grey bands provide a visual guide to compare the performance to: Kaiming-initialised networks when trained with 100,000 (top) and 50,000 (middle) images, and transfer learning using 100,000 images in the classifier-only case (bottom)



from 0.836 ± 0.004 to 0.896 ± 0.002 compared to training the network from scratch with 100k events. The validation F1 score is shown to demonstrate that the network was able to generalise well to the test sample.

Figure 4 shows the distribution of F1 scores from each of the 25 trained networks in the ensemble for the Kaiming-initialised (black) and the transfer learning all weights (red) cases, when trained using 100k interactions. The higher stability of the transfer learning case is shown clearly by the narrower distribution of F1 scores.

Figure 5 shows the class accuracy¹ for the three classes: $CC \nu_\mu$, $CC \nu_e$ and NC, for the transfer learning all weights and Kaiming-initialised networks. The accuracy for each class in the transfer learning case exceeds the corresponding class performance using the Kaiming initialisation. It demonstrates that the improvements from transfer learning come from improvement in all three classes.

¹ Typically called efficiency in high energy physics contexts.

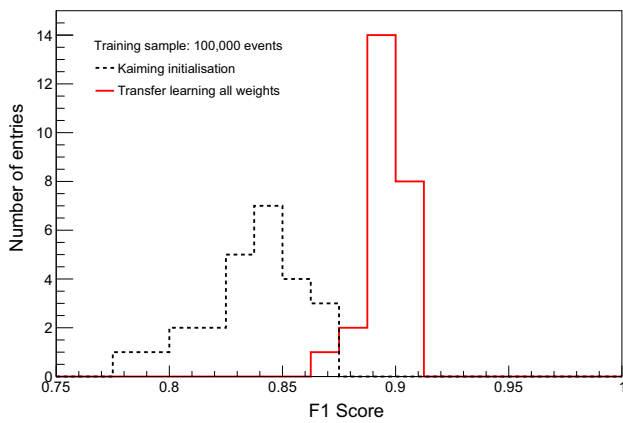


Fig. 4 F1 score distributions from the ensembles of 25 trained networks using the transfer learning all weights method (red) and the Kaiming-initialised networks (black). In both cases the networks were trained using the same 100,000 events

Table 3 shows two example confusion matrices from the Kaiming-initialised and transfer learning all weights networks trained on 100k images. In both cases, the network with the best F1 score of the 25 networks in the ensemble was chosen for presentation. The diagonal terms in these matrices show the number of correctly classified events, and the off-diagonal terms show the number of events wrongly classified as either of the other true classes. It can be seen that the transfer learning all weights network shows more correctly classified events and fewer incorrectly classified events for each of the true classes.

Table 3 Confusion matrices showing the number of correctly and incorrectly classified events for each true class. Results are shown for the network with the highest F1 score in the ensembles with 100,000 training images for Kaiming (top) and transfer learning all weights (bottom) cases

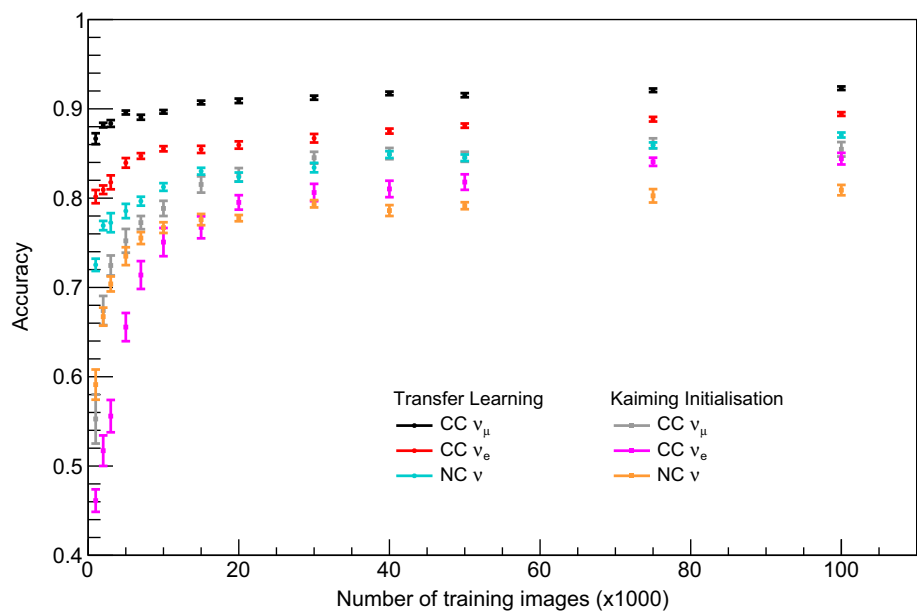
		Predicted		
		CC ν_μ	CC ν_e	NC
True	CC ν_μ	5839	386	487
	CC ν_e	276	5586	712
	NC	337	624	5753
Kaiming initialisation				
		Predicted		
		CC ν_μ	CC ν_e	NC
True	CC ν_μ	6262	163	287
	CC ν_e	111	6160	303
	NC	228	367	6119
Transfer learning all weights				

4.5 Comparison of classification bias with neutrino energy

To compare potential biases in classification performance as a function of true neutrino energy we consider the most performant networks, according to F1 score, from the ensemble of 25 networks trained with 100k events for each of the Kaiming-initialised and transfer learning all weights cases.

Classification accuracy (the fraction of correctly classified events for a given true interaction type) on the 20k event test sample is shown in Fig. 6. It is evident that the classification performance does vary with energy, and the pattern of variation is similar for both the Kaiming-initialised network

Fig. 5 The accuracy for the three classes (CC ν_μ , CC ν_e and NC) for the transfer learning all weights and Kaiming-initialised networks. The error bars show the uncertainty on the mean accuracy from the ensembles of 25 trained networks for each point



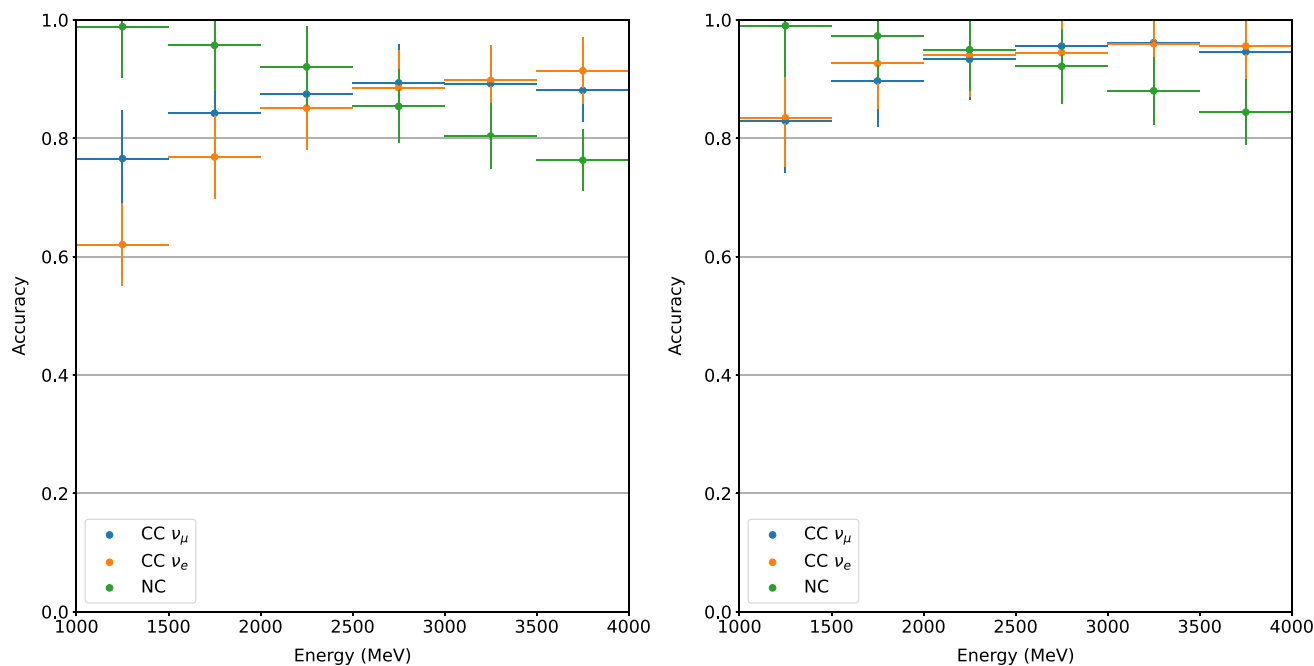


Fig. 6 The classification accuracy for each interaction type as a function of the true neutrino energy for the Kaiming-initialised (left) and transfer learning all weights (right) cases for networks trained with 100,000 images

and the transfer learned network. Performance in the two charged-current classes is reduced at lower energies, but the magnitude of the bias is notably less in the transfer learned case, with charged-current performance more nearly equivalent in each energy bin. For neutral current interactions we see a reduction in performance as energy increases, but once again the network trained via transfer learning shows less bias.

4.6 Comparison of classification bias with training sample size

Given one of the potential benefits of transfer learning is the ability to train on fewer events, it is worthwhile to explore how the number of events affects classification performance at the level of interaction type. To compare potential biases in classification performance as a function of the training sample size we consider the most performant network within the ensemble of transfer learned networks for each size of training sample.

First, it can be observed in Fig. 7 that the same pattern of behaviour is evident across all sample sizes, that is, charged-current classification accuracy improves as energy increases, while neutral current classification accuracy is reduced. However, it is also evident that the number of events introduces its own biases for the lowest training sample sizes. In particular, it can be seen that the magnitude of the reduction in classification accuracy for sample sizes below approximately 30k events is sensitive to the particular true interac-

tion type. Above this sample size, the improvement in classification accuracy is typically small, and similar for each interaction type.

Therefore, while the overall F1 score for transfer learning when training with only 7k events was broadly equivalent to that of the 100k event sample Kaiming-initialised network, it is clear that to achieve more balanced classification accuracy across interaction types and over a wide range of energies one should be cautious about pushing training sample size reduction too far without considering mitigating steps (for example, re-balancing the representation of each interaction type in the training sample). Though not presented here, such biases are, of course, also evident in the Kaiming-initialised networks, so this is not a unique feature of transfer learning.

5 Discussion

A systematic study comparing the performance of event classification in a DUNE-like LArTPC detector using standard CNN training methods and transfer learning was performed. It was found that the first convolutional layer of the ImageNet pre-trained ResNet18 extracted generic features that were applicable to LArTPC event images.

We have demonstrated that transfer learning can significantly outperform training randomly-initialised CNNs in the context of classifying neutrino interactions. We have also demonstrated that transfer learned networks exhibit reduced biases relative to networks trained from randomly initialised

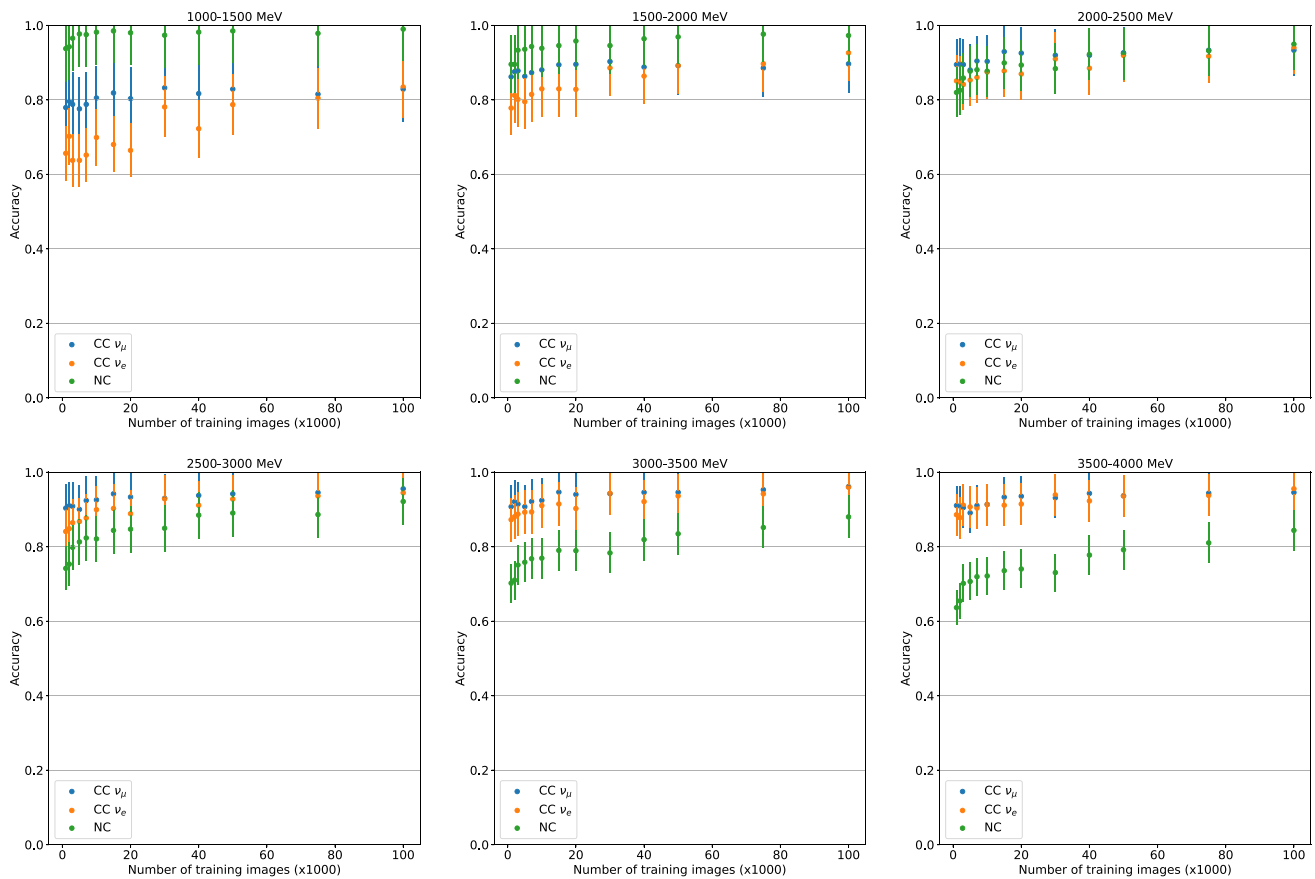


Fig. 7 The classification accuracy for each interaction type as a function of the training sample size for transfer learning all weights cases

weights. Fine tuning a pre-trained ResNet18 with only 7k images gave a better F1 score than a Kaiming-initialised ResNet18 trained on 100k images, though evidence of classification biases at such low sample sizes, whatever the training method, indicates caution is required in the use of very small samples. The results presented here demonstrate the potential of the transfer learning method as a way to obtain very good CNN performance with a relatively small number of training images.

Acknowledgements The authors thank Prof. A. Aurisano for the provision of computing resources at the University of Cincinnati, Dr S. Alonso Monsalve for his helpful comments on the manuscript, and Dr A. Basharina-Freshville for her work on a previous incarnation of the detector simulation.

Data availability This manuscript has no associated data or the data will not be deposited. [Authors' comment: Code to train the ensemble models analysed in this manuscript is available from the corresponding authors upon reasonable request.]

Declarations

Conflict of interest No funding was received for conducting this study, and the authors have no competing interests to declare that are relevant to the content of this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Funded by SCOAP³. SCOAP³ supports the goals of the International Year of Basic Sciences for Sustainable Development.

References

1. F. Psihas, M. Groh, C. Tunnell, K. Warburton, A review on machine learning for neutrino experiments. *Int. J. Mod. Phys. A* **35**(33), 2043005 (2020). <https://doi.org/10.1142/s0217751x20430058>
2. A. Radovic, M. Williams, D. Rousseau, M. Kagan, D. Bonaccorsi, A. Himmel, A. Aurisano, K. Terao, T. Wongjirad, Machine learning at the energy and intensity frontiers of particle physics. *Nature* **560**(7716), 41–48 (2018). <https://doi.org/10.1038/s41586-018-0361-2>

3. B. Abi et al., Neutrino interaction classification with a convolutional neural network in the DUNE far detector. *Phys. Rev. D* **102**(9), 092003 (2020). <https://doi.org/10.1103/PhysRevD.102.092003>. [arXiv:2006.15052](https://arxiv.org/abs/2006.15052) [physics.ins-det]
4. S.A. Monsalve, Novel usage of deep learning and high-performance computing in long-baseline neutrino oscillation experiments. Ph.D. thesis, Universidad Carlos III de Madrid (2020). <https://cds.cern.ch/record/2751646>
5. K. Bloom, V. Boisvert, D. Britzger, M. Buuck, A. Eichhorn, M. Headley, K. Lohwasser, P. Merkel, Climate impacts of particle physics (2022). [arXiv:2203.12389](https://arxiv.org/abs/2203.12389)
6. P. Calafiura, D. Rousseau, K. Terao, *Artificial Intelligence for High Energy Physics* (World Scientific, Singapore, 2022). <https://doi.org/10.1142/12200>
7. L. Dominé, K. Terao, Scalable deep convolutional neural networks for sparse, locally dense liquid argon time projection chamber data. *Phys. Rev. D* (2020). <https://doi.org/10.1103/physrevd.102.012005>
8. 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. Original in Croatian
9. S. Bozinovski, Reminder of the first paper on transfer learning in neural networks, 1976. *Informatica* **44**, 291–302 (2020)
10. F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, Q. He, A comprehensive survey on transfer learning. *Proc. IEEE* (2020). <https://doi.org/10.1109/JPROC.2020.3004555>
11. M.P. Kuchera, R. Ramanujan, J.Z. Taylor, R.R. Strauss, D. Bazin, J. Bradt, R. Chen, Machine learning methods for track classification in the AT-TPC. *Nucl. Instrum. Methods A* **940**, 156–167 (2019). <https://doi.org/10.1016/j.nima.2019.05.097>. [arXiv:1810.10350](https://arxiv.org/abs/1810.10350) [cs.CV]
12. M. Alam, et al., GENIE Production Release 2.10.0 (2015). <https://doi.org/10.48550/arXiv.1512.06882>. [arXiv:1512.06882](https://arxiv.org/abs/1512.06882) [hep-ph]
13. B. Abi et al., Long-baseline neutrino oscillation physics potential of the DUNE experiment. *Eur. Phys. J. C* **80**(10), 978 (2020). <https://doi.org/10.1140/epjc/s10052-020-08456-z>. [arXiv:2006.16043](https://arxiv.org/abs/2006.16043) [hep-ex]
14. S. Agostinelli et al., GEANT4: a simulation toolkit. *Nucl. Instrum. Methods A* **506**, 250–303 (2003). [https://doi.org/10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8)
15. B. Abi, et al., Deep Underground Neutrino Experiment (DUNE), Far Detector Technical Design Report, Volume IV: Far Detector Single-phase Technology. *JINST* **15**(08), 08010 (2020). <https://doi.org/10.1088/1748-0221/15/08/T08010>. [arXiv:2002.03010](https://arxiv.org/abs/2002.03010) [physics.ins-det]
16. A. Paszke, et al., Pytorch: an imperative style, high-performance deep learning library. In: Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., Garnett, R. (eds.) *Advances in neural information processing systems* 32, p. 8024–8035. Curran Associates, Inc., Vancouver, Canada (2019). <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>
17. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition. *CoRR* (2015). [arXiv:1512.03385](https://arxiv.org/abs/1512.03385)
18. C.J. Van Rijsbergen, *Information Retrieval*, 2nd edn. (Butterworths, London, 1979). <http://www.dcs.gla.ac.uk/Keith/Preface.html>
19. K. He, X. Zhang, S. Ren, J. Sun, Delving deep into rectifiers: surpassing human-level performance on ImageNet classification (2015). [arXiv:1502.01852](https://arxiv.org/abs/1502.01852)
20. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, ImageNet: a large-scale hierarchical image database. In: *CVPR09* (2009)
21. M.D. Zeiler, R. Fergus, Visualizing and understanding convolutional networks (2013). [arXiv:1311.2901](https://arxiv.org/abs/1311.2901)