Published for SISSA by Springer

Received: *September 24, 2020*
Revised: *November 20, 2020*
Accepted: *December 3, 2020*
Published: *January 28, 2021*

Demonstration of background rejection using deep convolutional neural networks in the NEXT experiment

The NEXT collaboration

M. Kekic,[u,s] C. Adams,[b] K. Woodruff,[c] J. Renner,[u,s] E. Church,[t] M. Del Tutto,[e] J.A. Hernando Morata,[u] J.J. Gómez-Cadenas,[p,i,1] V. Álvarez,[v] L. Arazi,[f] I.J. Arnquist,[t] C.D.R. Azevedo,[d] K. Bailey,[b] F. Ballester,[v] J.M. Benlloch-Rodríguez,[p,s] F.I.G.M. Borges,[n] N. Byrnes,[c] S. Cárcel,[s] J.V. Carrión,[s] S. Cebrián,[w] C.A.N. Conde,[n] T. Contreras,[k] G. Díaz,[u] J. Díaz,[s] M. Diesburg,[e] J. Escada,[n] R. Esteve,[v] R. Felkai,[f,g,s] A.F.M. Fernandes,[m] L.M.P. Fernandes,[m] P. Ferrario,[p,i] A.L. Ferreira,[d] E.D.C. Freitas,[m] J. Generowicz,[p] S. Ghosh,[k] A. Goldschmidt,[h] D. González-Díaz,[u] R. Guenette,[k] R.M. Gutiérrez,[j] J. Haefner,[k] K. Hafidi,[b] J. Hauptman,[a] C.A.O. Henriques,[m] P. Herrero,[p] V. Herrero,[v] Y. Ifergan,[f,g] B.J.P. Jones,[c] L. Labarga,[r] A. Laing,[c] P. Lebrun,[e] N. López-March,[v,s] M. Losada,[j] R.D.P. Mano,[m] J. Martín-Albo,[s] A. Martínez,[s,p] G. Martínez-Lema,[s,u,2] M. Martínez-Vara,[s] A.D. McDonald,[c] Z.-E. Meziani[b] F. Monrabal,[p,i] C.M.B. Monteiro,[m] F.J. Mora,[v] J. Muñoz Vidal,[s,p] P. Novella,[s] D.R. Nygren,[c,a] B. Palmeiro,[u,s] A. Para,[e] J. Pérez,[l] M. Querol,[s] A.B. Redwine,[f] L. Ripoll,[q] Y. Rodríguez García,[j] J. Rodríguez,[v] L. Rogers,[c] B. Romeo,[p,l] C. Romo-Luque,[s] F.P. Santos,[n] J.M.F. dos Santos,[m] A. Simón,[f] C. Sofka,[o,3] M. Sorel,[s] T. Stiegler,[o] J.F. Toledo,[v] J. Torrent,[p] A. Usón,[s] J.F.C.A. Veloso,[d] R. Webb,[o] R. Weiss-Babai,[f,4] J.T. White[o,5] and N. Yahlali[s]

[a] *Department of Physics and Astronomy, Iowa State University, 12 Physics Hall, Ames, IA 50011-3160, U.S.A.*
[b] *Argonne National Laboratory, Argonne, IL 60439, U.S.A.*
[c] *Department of Physics, University of Texas at Arlington, Arlington, TX 76019, U.S.A.*

[1] NEXT Co-spokesperson.
[2] Now at Weizmann Institute of Science, Israel.
[3] Now at University of Texas at Austin, U.S.A.
[4] On leave from Soreq Nuclear Research Center, Yavneh, Israel.
[5] Deceased.

Open Access, © The Authors.
Article funded by SCOAP[3].

https://doi.org/10.1007/JHEP01(2021)189

JHEP01(2021)189

[d] Institute of Nanostructures, Nanomodelling and Nanofabrication (i3N), Universidade de Aveiro, Campus de Santiago, Aveiro 3810-193, Portugal
[e] Fermi National Accelerator Laboratory, Batavia, IL 60510, U.S.A.
[f] Nuclear Engineering Unit, Faculty of Engineering Sciences, Ben-Gurion University of the Negev, P.O.B. 653, Beer-Sheva 8410501, Israel
[g] Nuclear Research Center Negev, Beer-Sheva 84190, Israel
[h] Lawrence Berkeley National Laboratory (LBNL), 1 Cyclotron Road, Berkeley, CA 94720, U.S.A.
[i] Ikerbasque, Basque Foundation for Science, Bilbao E-48013, Spain
[j] Centro de Investigación en Ciencias Básicas y Aplicadas, Universidad Antonio Nariño, Sede Circunvalar, Carretera 3 Este No. 47 A-15, Bogotá, Colombia
[k] Department of Physics, Harvard University, Cambridge, MA 02138, U.S.A.
[l] Laboratorio Subterráneo de Canfranc, Paseo de los Ayerbe s/n, Canfranc Estación E-22880, Spain
[m] LIBPhys, Physics Department, University of Coimbra, Rua Larga, Coimbra 3004-516, Portugal
[n] LIP, Department of Physics, University of Coimbra, Coimbra 3004-516, Portugal
[o] Department of Physics and Astronomy, Texas A&M University, College Station, TX 77843-4242, U.S.A.
[p] Donostia International Physics Center (DIPC), Paseo Manuel Lardizabal, 4, Donostia-San Sebastian E-20018, Spain
[q] Escola Politècnica Superior, Universitat de Girona, Av. Montilivi, s/n, Girona E-17071, Spain
[r] Departamento de Física Teórica, Universidad Autónoma de Madrid, Campus de Cantoblanco, Madrid E-28049, Spain
[s] Instituto de Física Corpuscular (IFIC), CSIC & Universitat de València, Calle Catedrático José Beltrán, 2, Paterna E-46980, Spain
[t] Pacific Northwest National Laboratory (PNNL), Richland, WA 99352, U.S.A.
[u] Instituto Gallego de Física de Altas Energías, Universidad de Santiago de Compostela, Campus sur, Rúa Xosé María Suárez Núñez, s/n, Santiago de Compostela E-15782, Spain
[v] Instituto de Instrumentación para Imagen Molecular (I3M), Centro Mixto CSIC, Universitat Politècnica de València, Camino de Vera s/n, Valencia E-46022, Spain
[w] Centro de Astropartículas y Física de Altas Energías (CAPA), Universidad de Zaragoza, Calle Pedro Cerbuna, 12, Zaragoza E-50009, Spain

E-mail: marija.kekic@usc.es

Abstract: Convolutional neural networks (CNNs) are widely used state-of-the-art computer vision tools that are becoming increasingly popular in high-energy physics. In this paper, we attempt to understand the potential of CNNs for event classification in the NEXT experiment, which will search for neutrinoless double-beta decay in ^{136}Xe. To do so, we demonstrate the usage of CNNs for the identification of electron-positron pair production events, which exhibit a topology similar to that of a neutrinoless double-beta decay

JHEP01(2021)189

event. These events were produced in the NEXT-White high-pressure xenon TPC using 2.6 MeV gamma rays from a ^{228}Th calibration source. We train a network on Monte Carlo-simulated events and show that, by applying on-the-fly data augmentation, the network can be made robust against differences between simulation and data. The use of CNNs offers significant improvement in signal efficiency and background rejection when compared to previous non-CNN-based analyses.

Keywords: Dark Matter and Double Beta Decay (experiments)

ArXiv ePrint: 2009.10783

JHEP01(2021)189

Contents

1 Introduction **1**

2 Topological signature **2**

3 Data acquisition and analysis **4**
3.1 The NEXT-White TPC 4
3.2 Event reconstruction 5

4 Convolutional neural network analysis **6**
4.1 Data preparation 6
4.2 Network architecture 8
4.3 Training procedure 11
4.4 Evaluation on data 13

5 Conclusions **17**

1 Introduction

Machine learning techniques have recently captured the interest of researchers in various scientific fields, including particle physics, and are now being employed in search of improved solutions to a variety of problems. In this study, we show that deep convolutional neural networks (CNNs) trained on Monte Carlo simulation can be used to classify, to a high degree of accuracy, events containing particular topologies of ionization tracks acquired from a high-pressure xenon (HPXe) time projection chamber (TPC). As CNNs trained on simulation are known to be difficult to apply directly to data due to the challenges associated with producing a Monte Carlo that perfectly matches experiment, we also present methods for extending the domain of application of a CNN trained on simulated events to include real events. We claim that our use of these methods in adapting CNNs to the experimental domain and verifying their performance is novel to the use of CNNs in the field.

Event classification is of critical importance in experiments searching for rare physics, as the successful rejection of background events can lead to significant improvements in overall sensitivity. The NEXT (Neutrino Experiment with a Xenon TPC) experiment is searching for neutrinoless double-beta decay ($0\nu\beta\beta$) in ^{136}Xe at the Laboratorio Subterráneo de Canfranc (LSC) in Spain. In the ongoing first phase of the experiment, the 5 kg-scale TPC NEXT-White [1] has demonstrated excellent energy resolution [2] and the ability to reconstruct high-energy (of the order of 2 MeV) ionization tracks and distinguish between the topological signatures of two-electron and one-electron tracks [3]. It has also been used to perform a detailed measurement of the background distribution and is expected to be capable of measuring the $2\nu\beta\beta$ mode in ^{136}Xe with 3.5σ sensitivity after 1 year of data-taking [4]. The next phase of the experiment, the 100 kg-scale detector NEXT-100, will search for the $0\nu\beta\beta$ mode at $Q_{\beta\beta}$, around 2.5 MeV. New techniques such as CNNs, which analyze the topology of an event near $Q_{\beta\beta}$ and aim to eliminate background events, are becoming more relevant and essential to reaching the best possible sensitivity.

Machine learning techniques have seen many recent applications in physics [5]. In neutrino physics in particular, CNNs have been applied to particle identification in sampling calorimeters in the NOvA experiment [6]. The MicroBooNE experiment has also employed CNNs for event classification and localization [7] and track segmentation [8] in liquid argon TPCs. IceCube has applied graph neural networks to perform neutrino event classification [9], and DayaBay identified antineutrino events in gadolinium-doped liquid scintillator detectors using CNNs and convolutional autoencoders [10]. Experiments searching for $0\nu\beta\beta$ decay have also employed CNNs: EXO has studied the use of CNNs to extract event energy and position from raw-waveform information in a liquid xenon TPC [11] and PandaX-III has performed simulation studies demonstrating the use of CNNs for background rejection in a HPXe gas TPC with a Micromegas-based readout [12]. Further simulation studies in HPXe TPCs with a charge readout scheme ("Topmetal") allowing for detailed 3D track reconstruction have also shown the potential of CNNs for background rejection in $0\nu\beta\beta$ searches [13]. NEXT has also presented an initial simulation study [14] of the use of CNNs for background rejection. In this study we show that CNNs can be applied to real NEXT data, using electron-positron pair production to generate events with a two-electron "$\beta\beta$-like" topology and studying how the energy distribution of such events changes when varying an acceptance cut on the classification prediction of a CNN.

The paper is organized as follows: section 2 describes the topological signature of a signal event. In section 3 the data acquisition and reconstruction is explained. A description of the CNN and training procedure, as well as evaluation on MC and data is given in section 4. Finally, conclusions are drawn in section 5.

2 Topological signature

In a fully-contained $0\nu\beta\beta$ event recorded by a HPXe TPC, two energetic electrons produce ionization tracks emanating from a common vertex. Though the fraction of energy $Q_{\beta\beta}$ carried by each individual electron may differ event-by-event, the general pattern observed is similar for the majority of events, and consists of an extended track capped on both ends

JHEP01(2021)189

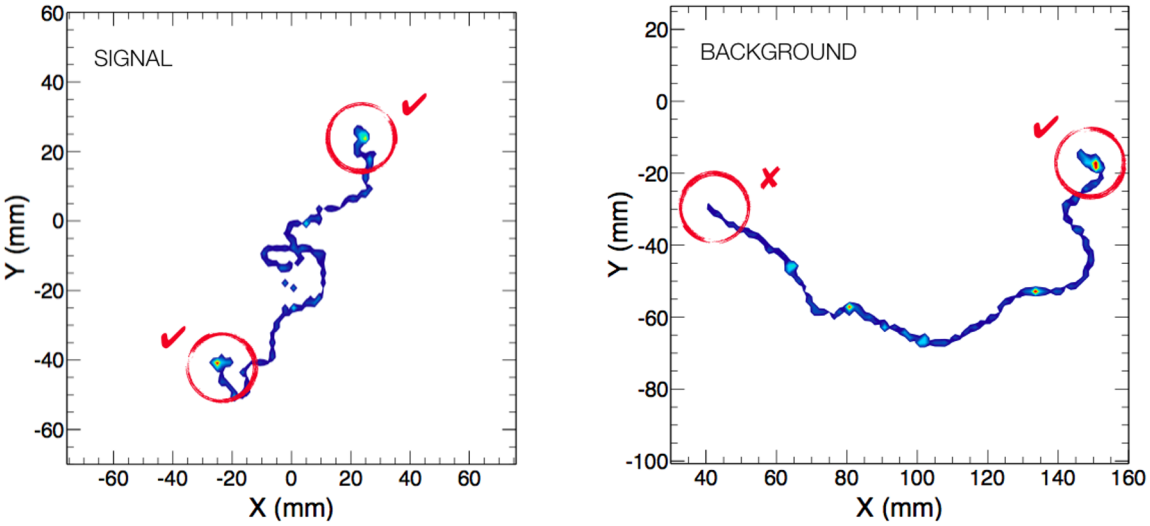


Figure 1. Energy depositions from trajectories in a Monte Carlo simulation of a $0\nu\beta\beta$ event, showing its distinct two-electron topological signature (left) compared with that of a single-electron event (right) of the same energy (figure from [15]).

by two "blobs", or regions of relatively higher ionization density. These regions are present due to the increase in stopping power experienced by electrons in xenon gas as they slow to lower energies. They provide a distinct signature for $0\nu\beta\beta$ decay, as measured tracks with similar energy produced by single electrons,[1] for example, photoelectric interactions of background gamma radiation, contain only one such "blob". The use of this signature, illustrated in figure 1, in performing background rejection is an essential part of the NEXT approach to maximizing sensitivity to $0\nu\beta\beta$ decay.

In order to demonstrate this approach experimentally, a reliable source of events with a similar topological signature is necessary. Electron-positron pair production by high energy gammas, followed by the subsequent escape from the active volume of the two 511 keV gamma rays produced in positron annihilation ("double-escape"), leaves a two-blob track formed by the electron and positron emitted from a common vertex, similar to the track that would be left by a $0\nu\beta\beta$ event. In this study, we use gamma rays of energy 2614.5 keV from ^{208}Tl (provided by a ^{228}Th calibration source, see figure 2) and observe the events in the double-escape peak at 1592 keV. This peak lies on top of a continuous background of single-electron tracks from Compton scattering of the calibration gamma rays and other background radiation. Experimentally, then, we have a sample containing $0\nu\beta\beta$-like events and background-like events. By evaluating these events with a Monte-Carlo-trained neural network and studying the resulting distribution of accepted events, we can demonstrate, using real data acquired with the NEXT-White TPC, the potential performance of such a network when employed in a $0\nu\beta\beta$ search. These results can be compared to a similar, non-CNN-based analysis published in [3].

[1]Events with multiple tracks are easier to reject simply by counting the number of isolated depositions.

JHEP01(2021)189

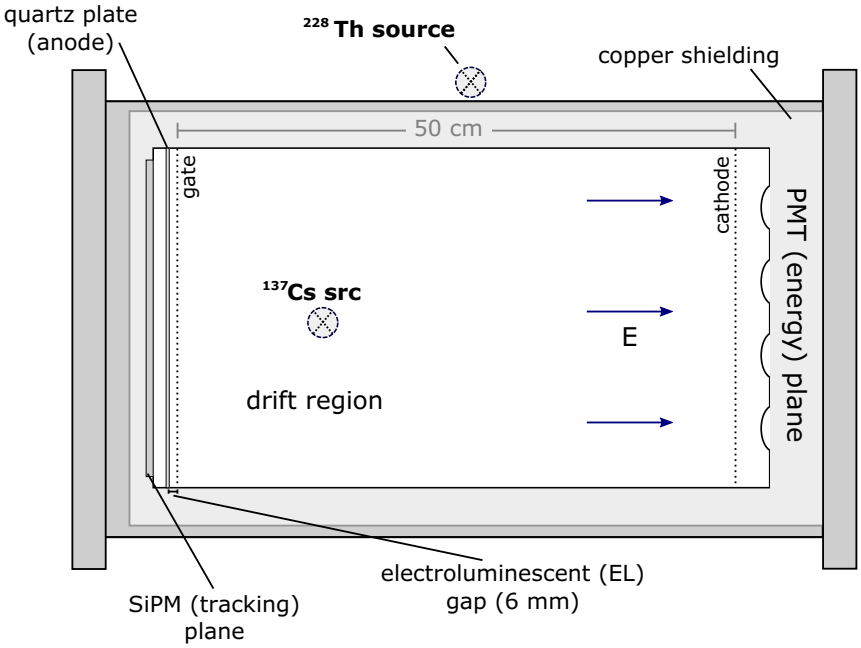


Figure 2. Schematic of the NEXT-White TPC, showing the positioning of the calibration sources (^{137}Cs and ^{228}Th) present during data acquisition for this study (figure derived from [2]).

3 Data acquisition and analysis

3.1 The NEXT-White TPC

The NEXT-White TPC measures both the primary scintillation and ionization produced by a charged particle traversing its active volume of high-pressure xenon gas. The main detector components are housed in a cylindrical stainless steel pressure vessel lined with copper shielding and include two planes of photosensors, one at each end, and several semi-transparent wire meshes to which voltages are applied, defining key regions of the detector (see figure 2). The two planes of photosensors are organized into an *energy plane*, containing 12 PMTs (photomultiplier tubes, Hamamatsu model R11410-10) behind the cathode, and a *tracking plane* containing a grid of 1792 SiPMs (silicon photomultipliers, SensL series-C, spaced at a 10 mm pitch) behind the anode. These sensors observe the scintillation produced in the active volume of the detector by ionizing radiation, including primary scintillation produced by excitations of the xenon atoms during the creation of the ionization track and secondary scintillation produced by *electroluminescence* (EL) of the ionization electrons. Note that in practice only the PMTs observe a consistently measurable primary scintillation signal, while EL is observed by both the PMTs and the SiPMs.

EL occurs after the electrons of the ionization track are drifted through the active region by an electric field (of order 400 V/cm) created by application of high voltage to the cathode (−30 kV) and gate (−7.6 kV) meshes and arrive at the EL gap, a narrow (6 mm) region defined by the gate mesh and a grounded quartz plate on which a conductive indium tin oxide (ITO) coating has been deposited. The large voltage drop over the narrow gap between the gate and the grounded plate creates an electric field high enough to accelerate

JHEP01(2021)189

the electrons to energies sufficient to excite the xenon without producing further ionization, allowing for better energy resolution compared to charge-avalanche detectors [16]. The subsequent decay of these excitations leads to EL scintillation, yielding of the order 500–1000 photons per electron traversing the EL gap. These photons, produced just in front of the tracking plane, cast a pattern of light on the SiPMs that can be used to reconstruct the (x, y) location of the ionization. The PMTs located in the energy plane on the opposite side of the detector see a more uniform distribution of light, including EL photons that have undergone a number of reflections in the detector, and record a greater total number of photons for a more precise measurement of the energy. The time difference between the observation of the primary scintillation (called S1) and secondary EL scintillation (called S2) gives the distance drifted by the ionization electrons before arriving at the EL region, corresponding to the z location at which this ionization was produced.

3.2 Event reconstruction

The data used in this study consisted of events with total energy near 1.6 MeV, including electron-positron events produced in pair production interactions from a 2.6 MeV gamma ray (see section 2) and background events, mostly due to Compton scattering of the same 2.6 MeV gamma rays.[2] The acquired signals for each event consisted of 12 PMT waveforms sampled at 25 ns intervals and 1792 SiPM waveforms sampled at 1 µs intervals for a total duration per read-out greater than the TPC maximum drift (approximately 500 microseconds). The ADC counts per unit of time in each waveform were converted to photoelectrons per unit time via conversion factors established by periodic calibration using LEDs installed inside the detector, a standard procedure in NEXT-White operation. The calibrations were performed by driving LEDs installed inside the vessel with short pulses and measuring the integrated ADC counts corresponding to a single photoelectron (pe).

The analysis of the acquired data was similar to that of [3]. The 12 PMT waveforms were summed, weighted by their calibrated gains, to produce a single waveform in which scintillation pulses were identified and classified as S1 or S2 according to their shape and location within the waveform. Events containing a single S1 pulse and at least one S2 pulse were selected, and for these events, the S2 information was used to reconstruct the ionization track. To do this, the S2 information was integrated into time bins of width 2 µs in both the PMTs and SiPMs. Note that to eliminate dark noise, SiPM samples with less than 1 pe were not included in the integration.

For each time bin, one or more energy depositions (or "*hits*") were reconstructed, and the pattern of signals observed on the SiPMs was used to determine the number of hits for a specific time bin and their corresponding (x, y) coordinates. A hit was assigned to the location of all SiPMs with an observed signal greater than a given threshold, and the total energy measured by the PMTs in that time bin was redistributed among the hits according to their relative SiPM signals.

The energy of each hit as measured by the PMTs was then corrected, hit-by-hit, by two multiplicative factors, one accounting for geometric variations in the light response in

[2]Environmental radioactivity was negligible compared to that of the source.

JHEP01(2021)189

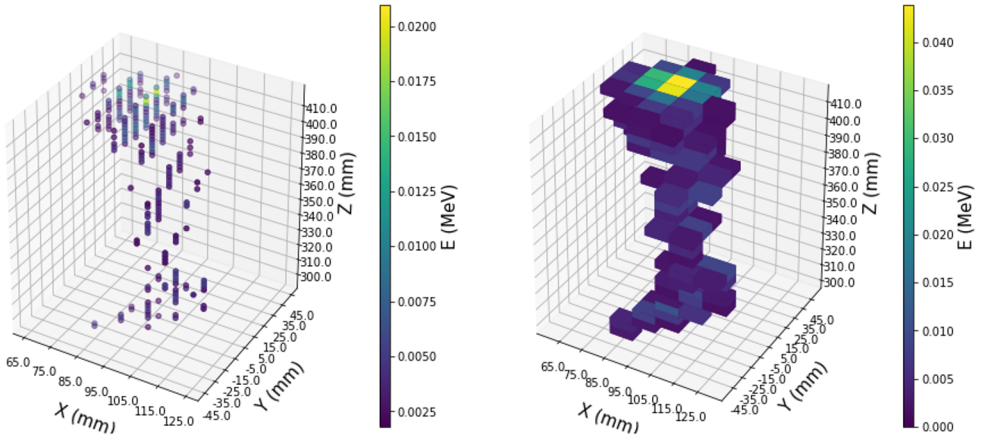


Figure 3. Reconstructed hits (left) and voxels (right) of a background Monte Carlo event. The volume within a tight bounding box encompassing the reconstructed hits is divided into $10 \times 10 \times 5$ mm^3 voxels to produced the voxelized track.

the EL plane and the other for electron attachment due to a finite electron lifetime in the gas. These correction factors were mapped out over the active volume by simultaneously acquiring events from decays of ^{83m}Kr, which was injected into the xenon gas and provided uniformly distributed point-like depositions of energy 41.5 keV [17]. The z-coordinate of each hit in the time bin was obtained from the time difference between S1 and S2 pulses, assuming an electron drift velocity of 0.91 mm/µs, as extracted from an analysis of the ^{83m}Kr events. A residual dependence of the event energy on the length of the event along the z-axis is observed, and a linear correction is performed to model this effect, which is not observed in simulation and remains to be fully understood. For details on this "axial length" effect, see [2].

The detector volume surrounding the reconstructed hits was then partitioned into 3D voxels of side length $10 \times 10 \times 5$ mm^3, and the energy of all hits that fell within each voxel was integrated. The X and Y dimensions of the individual voxels were chosen based on the 1 cm SiPM pitch, while the Z dimension was chosen to account for most of the longitudinal diffusion (1σ spread at maximum drift length is $\sim$ 2 m). The final voxelized track could then be considered in the neural-network-based topological analysis (see figure 3).

4 Convolutional neural network analysis

4.1 Data preparation

To generate the events used in training the neural network, a full Monte Carlo (MC) of the detector, including the pressure vessel, internal copper shielding, and sensor planes, was constructed using Nexus [18], a simulation package for NEXT based on Geant4 [19] (version `geant4.10.02.p01`). The ^{208}Th calibration source decay and the resulting interactions of the decay products were simulated by Geant4, up to and including the production of the ionization track. Events in the energy range of 1.4–1.8 MeV were selected, and the subsequent electron drift, diffusion, electroluminescence, photon detection, and electronic

JHEP01(2021)189

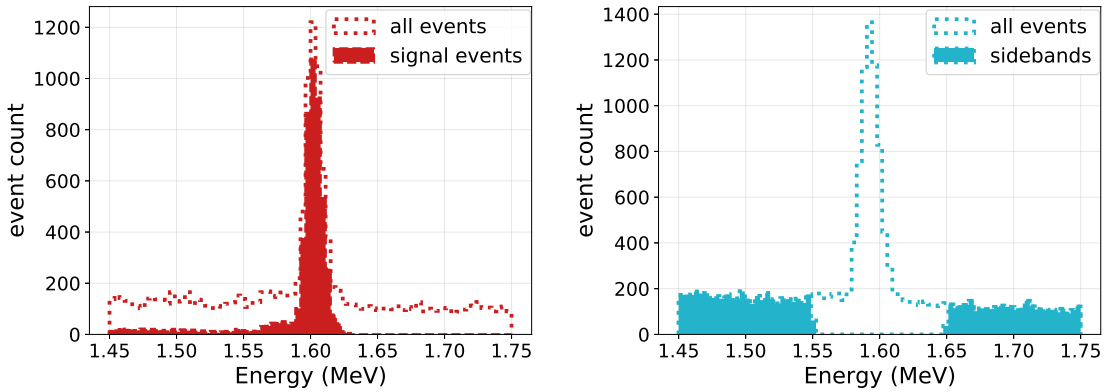


Figure 4. Left: energy distribution of all MC events (dashed line histogram) and of chosen signal events (solid histogram). Right: energy distribution of experimental data events showing selected sideband events. The sidebands are 100 keV in width, with each band starting 45 keV from either side of the double escape peak. The same procedure is also used to select the sidebands in MC.

readout processes were simulated outside of Geant4 to produce for each event a set of sensor waveforms corresponding to those acquired in NEXT-White. The analysis of data waveforms described in section 3.2 could then be applied to these MC waveforms to produce voxelized tracks (see figure 3). MC events that were fully contained in the active detector volume were used in the training set. To ensure the classification was done only based on the track topology, the energy of each voxel was scaled by the total event energy (the sum of voxel intensities for a given event was normalized to 1) such that the training data did not contain event energy information. Those events containing an electron and a positron registered in the MC true information, with no additional energy deposited by the two 511 keV gamma rays produced upon annihilation of the positron (i.e. a true "double-escape"), were tagged as "signal" events and all others were tagged as "background".

In [3], an additional single-track selection cut is made, and for a fair comparison with this previous result we also apply the same cut (obtained from the standard track reconstruction, for details see [3]) on test data only, for both MC and experimental data. As a reference, inside the peak energy range, the efficiency of the single-track cut was ~ 0.9 for signal events and ~ 0.7 for background events. For signal events, additional tracks appear either from physical processes such as bremsstrahlung, or from artificial splitting of the track due to imperfect reconstruction. The energy distribution of MC events with labeled signal events used for testing after the fiducial and single track selection cut is given in figure 4. A continuum of signal-like events below the 1.59 MeV peak energy belongs to gammas that Compton scatter before entering the detector, and undergo a pair conversion inside the detector active volume. The topology of those events is indistinguishable from the peak events since they only differ in the absolute reconstructed energy that is not taken into account during the training, and hence they are labeled as signal.

When applying a network trained on events from one domain (MC) to events from a different domain (data), the performance will depend on the similarity between those two domains. Known differences between MC and data in high-level variables, such as track length, have been observed in a previous topological analysis [3], calling into question the performance of the MC-trained CNN when applied to data. In this study, the classification

JHEP01(2021)189

task is focused on the double-escape peak, which is clearly visible and for which we understand the underlying physical process (pair production). In this case, we could attempt to design our network to obtain optimal classification results on the acquired data and in the energy range of interest (a method to evaluate the network on double-escape data is explained in section 4.4), but we could not argue that the same procedure would work in a $0\nu\beta\beta$ search for which we do not have a confirmed understanding of the underlying physics, nor would it be justified to make predictions on the same events used in optimizing the network.

Therefore, we develop a general paradigm (as described in section 4.3) that could be applied at $0\nu\beta\beta$ energies and, in evaluating the performance of the network on the data domain, uses events outside the energy range within which we intend to make predictions. Namely, before applying the CNN to the peak itself, we evaluate the performance on the peak sidebands (see figure 4), where the sample composition is known, and we expect the CNN predictions to be similar in data and MC. The underlying assumption is that the domain shift between MC and data is not correlated with the type of event, i.e. we expect that if a network is robust to MC/data differences on sidebands, it will be robust to MC/data differences in the peak region as well. In [3] it was shown that the track length difference between data and MC is consistent across a wide energy range, giving us confidence that the differences are indeed coming from the detector simulation and reconstruction (which should have the same effect on both signal and background events), rather than incorrectly simulated physical processes, justifying the sidebands-testing approach.

4.2 Network architecture

After the initial proposal of a deep convolutional architecture for image recognition [20], all modern architectures are a combination of a set of basic layers: convolutional, downsampling (pooling) and fully connected (dense) layers. They are described as follows:

- Convolution is performed by linearly applying *filters*, which are tensors of learnable parameters whose dimensions correspond to the spatial dimensions of the input image, and stacking the output of each filter. For example, in the case of a two-dimensional colored image, a typical filter would have size $5\times5\times3$ (the last dimension corresponds to the *feature*[3] dimension, which in the case of colored images represents the three RGB channels) and would slide[4] over the width and height of the input image, at each step applying a dot product over the third dimension and producing a 2D output. The process is repeated *number-of-filters* times and the 2D outputs are stacked into a final layer output.
- Pooling layers are used to reduce the spatial dimensions of the intermediate outputs, and usually a maximum or average value of the features inside a small window (typical size is 2) is used.
- Fully connected layers, the basic layer of shallow neural networks, perform a matrix multiplication of a 1D input vector with a matrix of parameters.

[3] In machine learning language, features are sets of numbers extracted from input data that are relevant to the task being learned.

[4] The step of the sliding is called *stride* and the usual values are 1–2.

JHEP01(2021)189

Other standard layers include dropout layers [21], where in each forward pass some number of features is randomly dropped during the training to ensure that the prediction is not dominated by some specific features, and normalization layers, in which the outputs of the intermediate layers are rescaled to avoid numerical instabilities. The most commonly employed normalization layer is the batch normalization layer [22].

A nonlinear activation function is applied[5] to the output of each convolutional and dense layer. The choices of the activation functions are many, but the standard ones are the rectified linear unit (ReLU) function for intermediate layers, and the Softmax function for the final output in the case of classification. Softmax is used to map the input vectors to the range $[0, 1]$, hence allowing the outputs of the network to be interpreted as probabilities.[6] The number and order of layers and the number of filters and their sizes should be adapted considering the complexity of the problem, the size of the training dataset and the available computational resources.

For the particular choice of network architecture in this study, we followed state-of-the-art practices in computer vision and deep neural networks, with special adaptations for the nature of our data. The most common building block for image analyzing networks is the *residual block* pioneered in [24], with our implementation pictured in figure 5b. This building block uses convolutional filters to learn a modification to its input layer rather than a direct transformation: $R(x) = x + f(x)$, where the residual block R is learning small changes (f) to an image. The network used here consisted of two initial convolutional layers, and a set of pre-activated residual block layers [25] (a particular variant of residual layers) followed by two consecutive dense layers with a dropout layer before each, as seen in figure 5a. The nonlinear activation employed is the standard ReLU preceded by a batch normalization layer in all but the fully connected layers.[7]

A major modification to our network from the residual networks in [24] is the use of three-dimensional convolutions as opposed to two-dimensional convolutions. Additionally, we employ fewer layers since our classification task with two output classes is simpler than the thousand-class challenges confronted by ImageNet [27], for which residual networks were originally developed.

In our network, the input dimensions were $40 \times 40 \times 110$ with each input corresponding to one voxel, therefore covering a volume of $40 \times 40 \times 55\,\mathrm{cm}^3$, essentially the entire active volume of the detector. The data are locally dense but globally sparse, as is common in high-energy physics and nuclear physics. In this dataset, the number of voxels with activity is typically $\mathcal{O}(100)$, compared to 176,000 voxel locations in the input space (0.05% pixel occupancy).

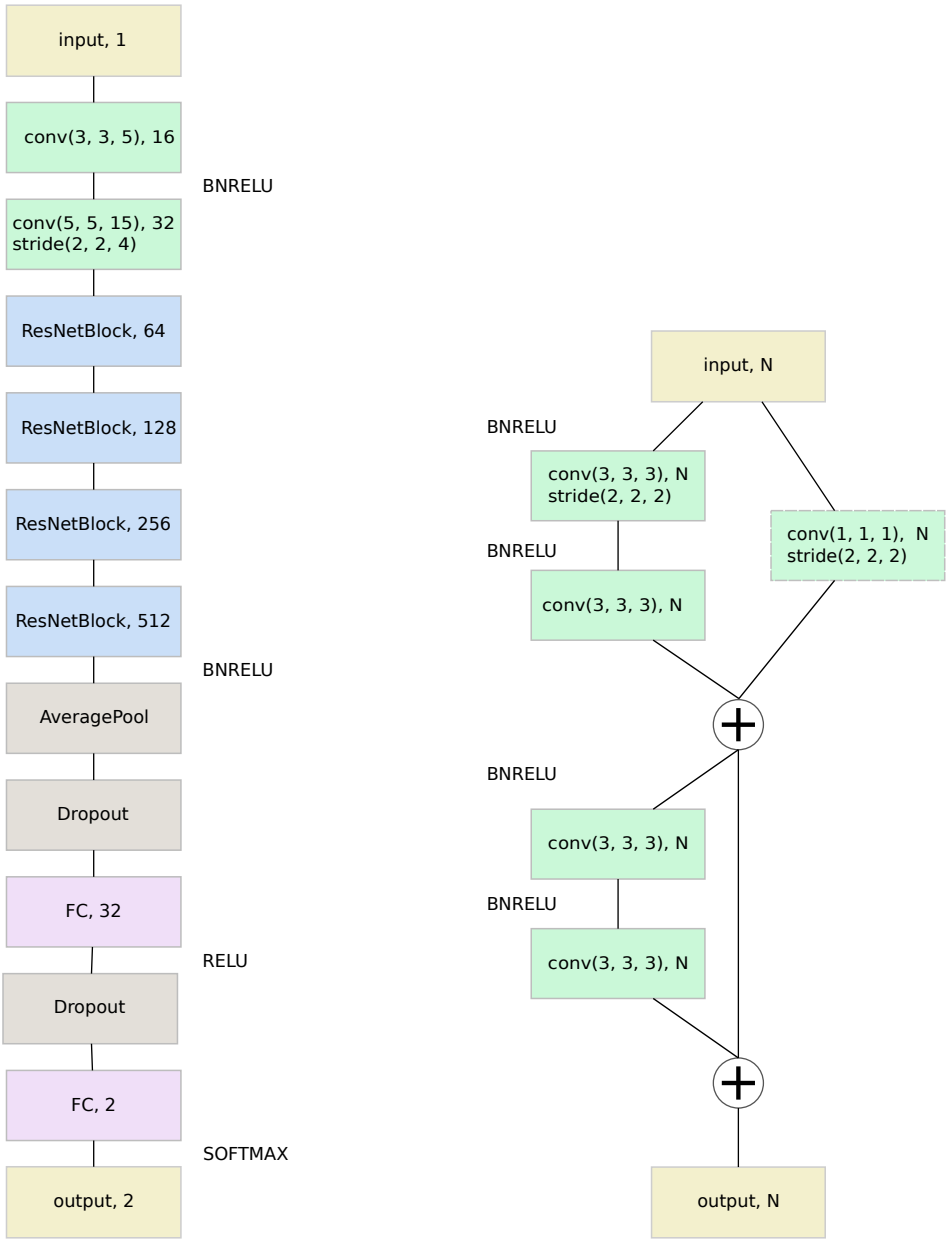
While traditional dense 3D convolutional neural networks will work on this dataset, we employed the Submanifold Sparse Convolutional Networks (SCN) framework [28], implemented in PyTorch. SCN is highly suitable for sparse input data, making the linear algebra far more computationally efficient than with non-sparse techniques.

[5] Without nonlinear activations, the whole network would simply collapse into a linear model.

[6] The predictions of deep networks should be scaled to truly represent probabilities (in the frequentist sense) [23].

[7] Batch normalization layers mixed with Dropout layers can lead to numerical instabilities [26].

JHEP01(2021)189



(a) ResNet architecture. **(b)** ResNet pre-activated block.

Figure 5. (a) Summary of the neural network architecture used in this analysis, with (b) details of each ResNetBlock architecture. The following notation is used: conv(fx, fy, fz), N represents a Convolutional 3D layer with filter size (fx, fy, fz) and N filters, with stride(sx, sy, sz) shown in the same cell if stride > 1 is applied. FC, N represents a fully connected layer with N output neurons. AveragePool is applied after the last ResNetBlock and Dropout with $p = 0.5$ is applied before each FC layer. Nonlinear activation throughout the network is ReLU, except for the final activation which is Softmax. BatchNorm normalization layer is applied before ReLU activation for all but the last ReLU layer.

JHEP01(2021)189

Such networks have already been used in high-energy physics analysis [29] and the main advantage of these types of network is that they occupy less memory and allow for larger input volumes and/or larger batch sizes,[8] as well as significantly faster training times. All of the results shown here were obtained using this framework, but we obtained similar results using the standard implementation of dense convolutions in Keras/TensorFlow.

We note, in particular, that the dataset used here from NEXT-White allows for the use of dense convolutional neural networks, but this will not be true in future detectors. Applications of this technology at larger scales, including NEXT-100 and beyond, will lead to a total number of voxels that scales as the volume of the detector, growing with a cubic power. Current and future AI accelerator hardware (GPUs, for example) will not be able to process the dense volumes of a ton-scale detector. However, the number of active voxels for an interesting physics event will likely stay the same as in NEXT-White, or increase modestly with improved resolution. Therefore, sparse convolutional networks will be scalable to even the largest high pressure TPCs in the NEXT program.

The main result of this paper is the achievement of reliable results from CNN-based tools that operate on real data, rather than working towards small gains in improvement on simulation with a neural architecture search. As will be shown later, the training has to be constrained to prevent domain overfitting, hence some improvement on simulated data could be achieved with different architectures but would not necessarily be reflected in experimental data.

4.3 Training procedure

Optimizing learnable parameters of convolutional and fully connected layers to achieve the desired predictions is called training, and is done by minimizing the disagreement (*loss*) between the true labels and the predicted ones.[9] The most common loss function employed in classification problems with neural networks is the cross entropy loss $H(p,q) = -\sum_x p(x) \log q(x)$, where $p(x)$ are the true labels and $q(x)$ the predicted ones, calculated over all input images of one batch.

During the training, the performance of the network is tracked on the validation set, which is independent of the training set and is not used for network optimization. In this study, a total of about 500 k simulated fiducial events were used as a training set, of which 200 k were signal events, and an additional $\sim$ 30 k events were used as a validation sample with similar signal proportion. A batch size of 1024 was chosen, and the cross entropy loss was weighted according to the signal-to-background ratio of the entire dataset. To avoid overfitting, in addition to the use of dropout, L2 weight regularization (a term in the loss function that penalizes high values of network parameters) and on-the-fly data augmentation[10] [30] were employed. The augmentation includes translations, dilation or "zooming" (scaling all 3 axes independently), flipping in x and y, and varying SiPM charge

[8] Batch size is the number of images processed in one forward pass through the network. Larger batch sizes allow for faster training.

[9] This is a general paradigm behind *supervised* machine learning.

[10] On-the-fly means that the augmentation is done during the code execution and the augmented dataset is not stored on the disk.

JHEP01(2021)189

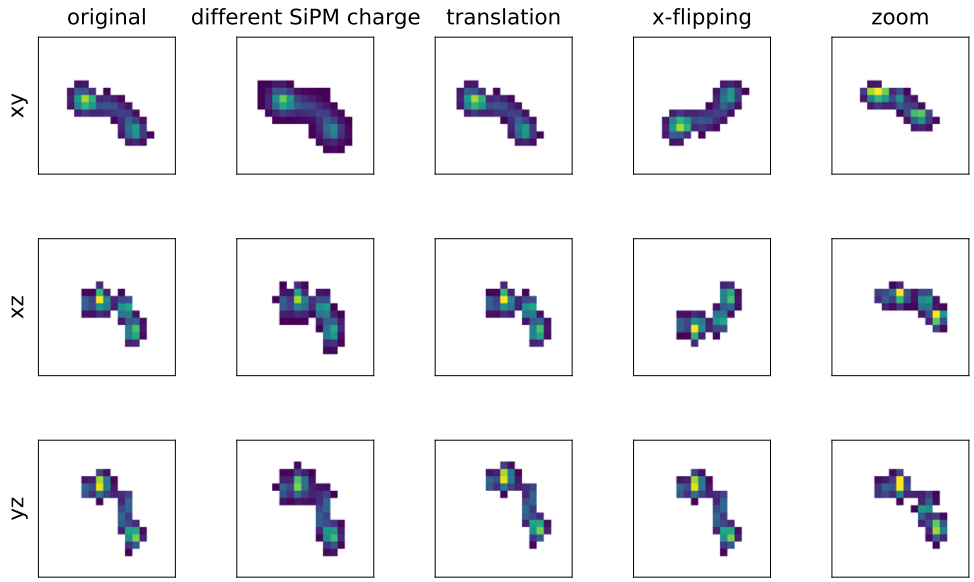


Figure 6. Example of on-the-fly data augmentation used during training on a selected signal event, projected on three planes for easier visualization.

cuts[11] as detailed in figure 6. We note that augmentation procedures used here are explicitly designed to be "label preserving" in that they do not change the single or double-blob nature of events, but do reduce the significance of differences in data/simulation.

As noted in section 4.1, since CNNs are highly nonlinear models, their application outside the training domain cannot be assumed to be reliable, and before applying the network to events in the peak we compare extracted features of MC and data events on the sidebands. It is common to consider convolutional layers as feature extractors (each one extracting higher level features), and consecutive dense layers as a classifier. The paradigm of features extractors followed by a classifier is not unique to machine learning approaches. For example, the previous analysis of NEXT event topology used graph methods to determine connected tracks and extract track properties such as length, endpoints and the energies around them (so-called *blobs*), which can be considered as a feature extractor. A classifier in that case was a cut on the obtained blob energies. Later on, the matching of MC and data features was done by rescaling MC features (for further explanation and justification of the procedure we refer the reader to [3]).

The features in CNN models do not necessarily have intuitive meaning and should rather be considered as a dimensionality reduction that preserves relevant information from the input image needed for the classification task. Hence, the choice of the layer to be considered as the last one of the feature extractor is somewhat arbitrary. We chose the output of the AveragePool layer (figure 5) as a representative feature vector, resulting in 512-dimension features. Since the features are not intuitive, any rescaling would not be justified and we are limited to simply tracking how MC features differ from those of

[11]For validation and testing, the SiPM cut was fixed at 20 photoelectrons, while in the augmentation it was allowed to vary ±10 around this value.

JHEP01(2021)189

data on the sidebands where the sample composition is known, and hence we expect similar distributions if the network is indeed robust to MC/data differences. To quantify similarity between multidimensional distributions we applied a two sample test; a test to determine whether independent random samples of R^d-valued random vectors are drawn from the same underlying distribution. We chose energy test statistics [31, 32], but the qualitative result should not depend on the choice of the metrics. The energy distance between two sets A, B is given by

$$\epsilon_{A,B} = \frac{2}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}\|x_i - y_j\| - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|x_i - x_j\| - \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\|y_i - y_j\| \tag{4.1}$$

where x_i, y_i are n, m samples drawn from the two sets. In [32], it was proven that this quantity is non-negative and equal to zero only if x_i and y_i are identically distributed, hence the energy distance is indeed a metric. The p-value, or probability of observing an equal or more extreme value than the measured value, for rejecting the null hypothesis (in this case, rejecting the hypothesis that MC and data features follow the same underlying distribution) can be calculated via the permutation test [33]. Namely, the nominal energy distance is computed, and the x_i and y_j are then divided into many (1000 in our case) possible arrangements of two groups of size n and m. The energy distance is computed again for each of these arrangements, each of which corresponds to one permutation. The p-value is given by the fraction of permutations in which the energy distance was larger than the nominal one.

The training and validation losses are given in figure 7 for the networks trained with and without data augmentation. The overfitting apparent in the case of training without augmentation is prompt and is manifested in the divergence of the validation and test losses, meaning that the network is beginning to memorize the training dataset and is not generalizing well. In figure 8 we show that the data augmentation also reduces the data/MC features distribution distance (eq. (4.1)), giving us more confidence that the performance on data will be similar to the performance on MC. As the distances are always calculated on MC and data events directly (without applying any data augmentation transformations), this technique does not directly correct MC but rather makes the model more robust to the data/MC differences. The final model is chosen by varying regularization parameters and selecting the training iteration step that gives minimal classification loss on the MC validation sample, ensuring that the corresponding p-value of energy test statistics is not larger than 5%.

4.4 Evaluation on data

In an ideal test of the trained network, we would have a data sample of only e^+e^- events at the energy of interest acquired from our detector, and another sample of single-electron events at the same energy. However, as we will always have background events, in particular due to Compton scattering of the high-energy gamma rays used in producing the e^+e^- events with the topology of interest, an exactly-labeled test set of detector data is impossible. Therefore we make an assumption about the characteristics of the energy spectrum

JHEP01(2021)189

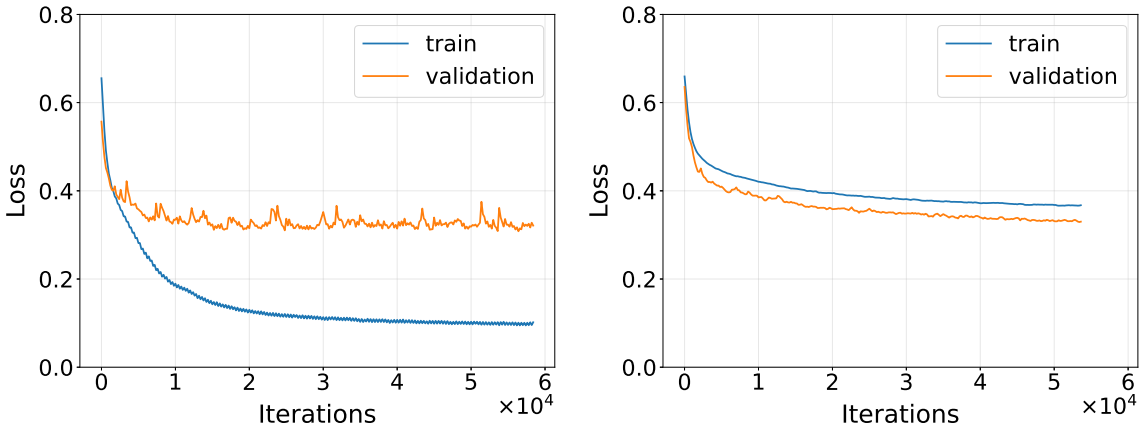


Figure 7. Training and validation losses without (left) and with (right) the application of data augmentation to the training set. Overfitting in the left-hand plot is visible only after 1000 iterations. As the augmentation procedure is only relevant to the training phase, it was not applied to the validation set. The ability of the network to make correct predictions is improved for events unaltered by data augmentation, which explains why the loss is higher for the training set than for the validation set in the right-hand plot.

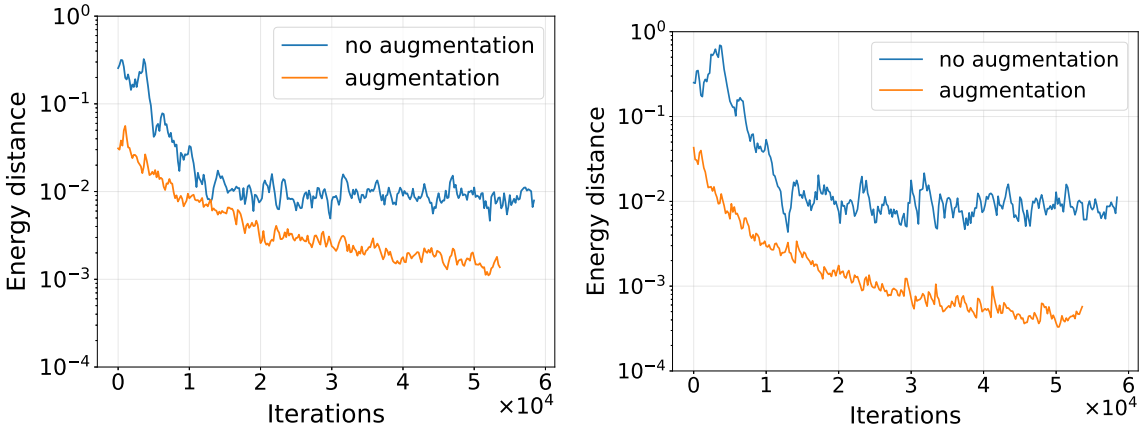


Figure 8. Energy distance between data and MC features during the training on the left sideband (left) and right sideband (right) for training with and without the augmentation. The corresponding p-value for the chosen model with augmentation at the chosen iteration step was ~ 0.1 (0.2) for the left (right) sideband.

near the energy of interest and attempt to extract the number of signal and background events present, following the procedure explained in [3].

First, we select only fiducial events passing a single-track cut as explained in section 4.1. Note that the single-track cut was not applied to the training set, but we do apply it to the test set to allow for exact comparison with the previous analysis. We then assume that the signal events produce a Gaussian peak (as indeed would be the case for events occurring at a precise energy), and that the background, consisting of Compton electrons, in the region of the peak can be characterized by an exponential distribution. The peak energy region is fixed to 1.570–1.615 MeV (as in [3]), a region that contains more than 99.5% of the

JHEP01(2021)189

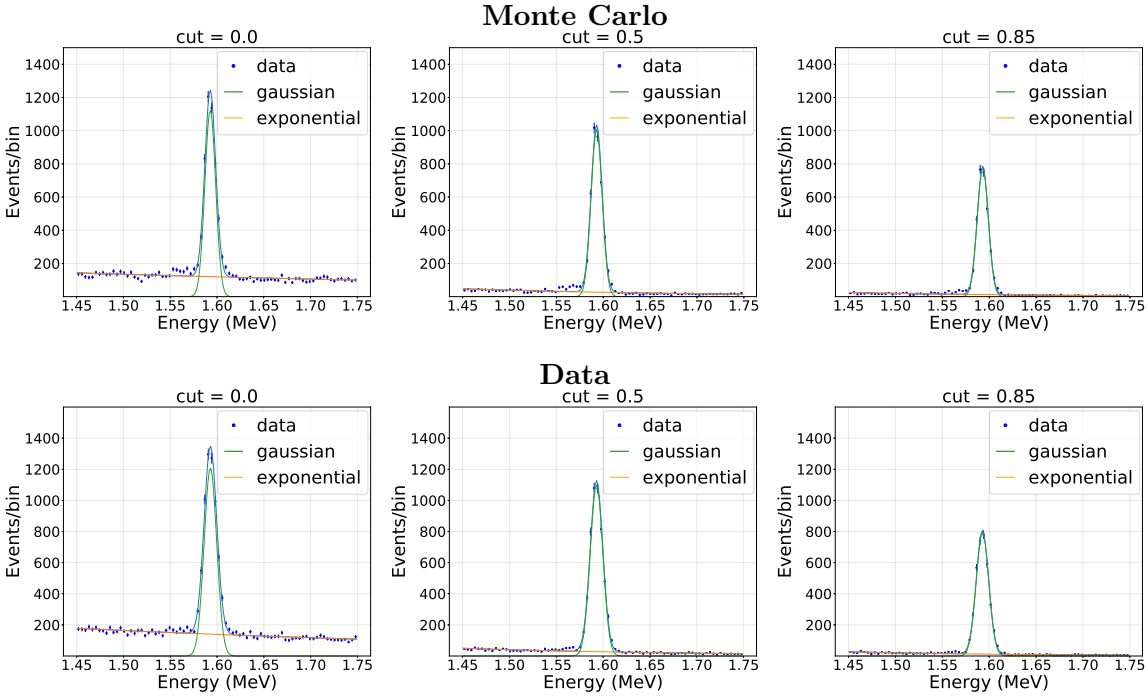


Figure 9. Fits of the energy spectra of Monte Carlo simulation (top) and data (bottom) near the double-escape peak of ^{208}Tl at 1592 keV (see text for details). The fits are shown using events passing a neural network classification cut of 0.0 (to obtain the total number of signal and background events), 0.5 (the cut usually used in binary classification problems) and 0.85 (the cut that was found to yield the optimal f.o.m.).

Gaussian peak for both data and MC. Then, we apply an unbinned fit of the sum of two curves (Gaussian + exponential) to the full energy spectrum in the larger energy range[12] 1.45–1.75 MeV in order to keep the fits stable, obtaining the parameters defining the two curves. Integrating over theoretical Gaussian and exponential curves in the peak energy range gives us the estimate of the initial number of signal events s_0 (from the Gaussian) and the initial number of background events b_0 (from the exponential). This procedure is then repeated using the spectra obtained from events with network classification greater than a varying threshold, in each case obtaining the number of accepted signal events s and accepted background events b.

Figure 9 illustrates this fit procedure for three different threshold values on the CNN prediction output using data from NEXT-White and a set of Monte Carlo simulated events that were not present in the training set. Varying the classification threshold traces out a curve in the space of signal acceptance s/s_0 vs. background rejection $1-b/b_0$ (see figure 10). To obtain optimal sensitivity in a $0\nu\beta\beta$ search, one must maximize the ratio of accepted signal to the square root of the rejected background [15], and therefore we also construct the figure of merit $F = s/\sqrt{b}$ for the various classification thresholds. We show for comparison

[12]Note the slightly narrower energy range compared to the one specified in section 4.1 which is used to pre-select MC events based on true energy deposited in the detector. This choice is made to remove artificial disturbances at the end points of the selected energy range.

JHEP01(2021)189

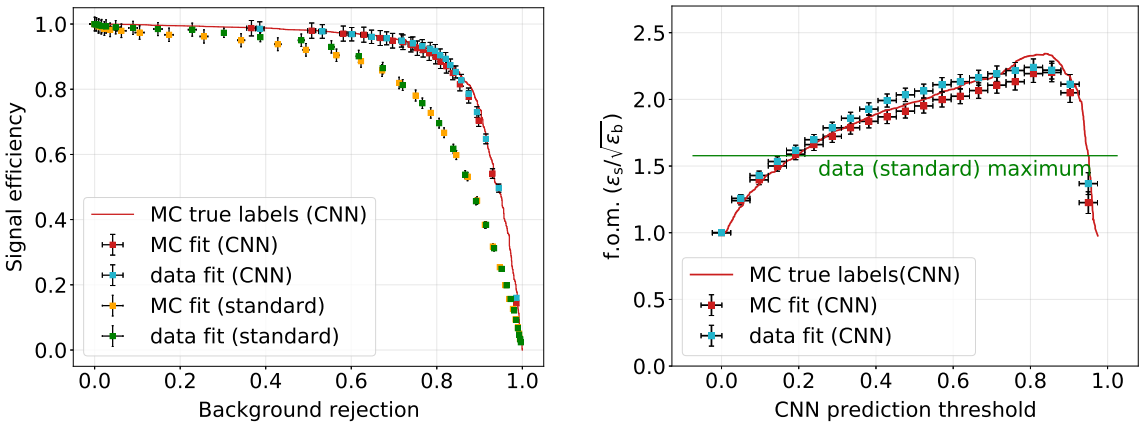


Figure 10. The signal acceptance vs. background rejection (left) and the figure of merit (right). The curves labeled "fit" are traced out by varying the neural network classification threshold and determining the fraction of accepted signal and rejected background using the fit procedure described in the text, while the MC true labels are obtained using MC labels as described in section 4.1.

the non-CNN-based result obtained in [3]. In Monte Carlo, we find a maximum figure of merit of $F = 2.20$ with signal acceptance $s/s_0 = 0.70$ and background rejection $1 - b/b_0 = 0.90$. In data, fixing the CNN cut to the one giving the best Monte Carlo figure of merit, we find $F = 2.21$, with signal acceptance $s/s_0 = 0.65$ and background rejection $1-b/b_0 = 0.91$. Compared to the previous result, the CNN analysis yields an improvement in the figure of merit of a factor $2.2/1.58 \sim 1.4$. Assuming that this factor stays the same at $0\nu\beta\beta$ energies, the improvement in the half-life measurement would be directly proportional to this factor, while the improvement in the measurement of $m_{\beta\beta}$ would be $\sqrt{1.4} \sim 1.16$. However, these factors should not be taken as exact since both traditional and CNN approaches depend on the detector specifics and reconstruction abilities, as well as the precision of the simulations.[13] The sensitivity predictions of the traditional event selection for the next two detectors planned can be found in [15, 34].

We note that in figure 10 there is excellent agreement between data and simulation when comparing the signal efficiency in this analysis at a fixed background rejection, but there is still a minor disagreement between the figure of merit for simulation and data as a function of prediction threshold. Several reasons account for this disagreement. First, the data-augmentation technique extends the domain of applicability of the neural networks trained solely on simulated data, but it does not account for all possible differences between the data and Monte Carlo events. For example, any effect that would redistribute the energy along the track is not covered by the transformations we employ in data-augmentation. We anticipate that many of the effects contributing to data/simulation disagreement, such as the axial length effect mentioned in section 3.2, will be understood and resolved in the future and will bring these minor residual differences even closer together. A smaller EL TPC built from the original hardware of the NEXT-DEMO prototype [35, 36] is currently operational and will provide data that can be used to study these effects in more detail.

[13] In the case of no domain discrepancy between experimental and simulated data, the amount of regularization could be relaxed and the performance would improve.

JHEP01(2021)189

Second, the fit procedure error could account for some of the differences in the figure of merit plot. Namely, modeling the energy distribution as the sum of a Gaussian signal and exponential background does not adequately account for the long left tail visible in figure 4. For low prediction threshold cuts, while the background acceptance is still high, these events are a minor effect, but as the threshold cut increases they become a larger portion of the left sideband during the fit procedure, leading to an underestimated signal efficiency when compared to the efficiency calculation on simulation obtained using the true underlying event type (the mismatch of red points and the continuous red line in figure 10). A different ratio of signal inside the left sidebands between data and MC could lead to a different figure of merit.

5 Conclusions

We have demonstrated the first data-based evaluation of track classification in HPXe TPCs with neural networks. The results confirm the potential of the method demonstrated in previous simulation-based studies and show that neural networks trained using a detailed Monte Carlo can be employed to make predictions on real data. The present results show that the background contamination can be reduced to approximately 10% while maintaining a signal efficiency of about 65%. In fact, these results are likely to be conservative, as this demonstration was performed at an energy of 1592 keV, while at the same pressure, tracks with energy $Q_{\beta\beta}$ are longer, and therefore their topological features should be more pronounced.

Furthermore, we have shown that, with the application of appropriate domain regularization techniques to the training set, our model performs similarly on detector data and simulation in the extraction of the signal events of interest.

Acknowledgments

This study used computing resources from Artemisa, co-funded by the European Union through the 2014-2020 FEDER Operative Programme of the Comunitat Valenciana, project DIFEDER/2018/048. This research used resources of the Argonne Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC02-06CH11357. The NEXT collaboration acknowledges support from the following agencies and institutions: Xunta de Galicia (Centro singularde investigación de Galicia accreditation 2019-2022), by European Union ERDF, and by the "María de Maeztu" Units of Excellence program MDM-2016-0692 and the Spanish Research State Agency"; the European Research Council (ERC) under the Advanced Grant 339787-NEXT; the European Union's Framework Programme for Research and Innovation Horizon 2020 (2014-2020) under the Grant Agreements No. 674896, 690575 and 740055; the Ministerio de Economía y Competitividad and the Ministerio de Ciencia, Innovación y Universidades of Spain under grants FIS2014-53371-C04, RTI2018-095979, the Severo Ochoa Program grants SEV-2014-0398 and CEX2018-000867-S; the GVA of Spain under grants PROMETEO/2016/120 and

JHEP01(2021)189

SEJI/2017/011; the Portuguese FCT under project PTDC/FIS- NUC/2525/2014 and under projects UID/FIS/04559/2020 to fund the activities of LIBPhys-UC; the U.S. Department of Energy under contracts number DE-AC02-07CH11359 (Fermi National Accelerator Laboratory), DE-FG02-13ER42020 (Texas A&M) and DE-SC0019223/DE SC0019054 (University of Texas at Arlington); and the University of Texas at Arlington. DGD acknowledges Ramon y Cajal program (Spain) under contract number RYC-2015 18820. JM-A acknowledges support from Fundación Bancaria "la Caixa" (ID 100010434), grant code LCF/BQ/PI19/11690012. We also warmly acknowledge the Laboratori Nazionali del Gran Sasso (LNGS) and the Dark Side collaboration for their help with TPB coating of various parts of the NEXT-White TPC. Finally, we are grateful to the Laboratorio Subterráneo de Canfranc for hosting and supporting the NEXT experiment.

Open Access. This article is distributed under the terms of the Creative Commons Attribution License (CC-BY 4.0), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

References

[1] NEXT collaboration, *The Next White (NEW) Detector*, 2018 *JINST* **13** P12010 [arXiv:1804.02409] [INSPIRE].

[2] NEXT collaboration, *Energy calibration of the NEXT-White detector with 1% resolution near $Q_{\beta\beta}$ of ^{136}Xe*, *JHEP* **10** (2019) 230 [arXiv:1905.13110] [INSPIRE].

[3] NEXT collaboration, *Demonstration of the event identification capabilities of the NEXT-White detector*, *JHEP* **10** (2019) 052 [arXiv:1905.13141] [INSPIRE].

[4] NEXT collaboration, *Radiogenic Backgrounds in the NEXT Double Beta Decay Experiment*, *JHEP* **10** (2019) 051 [arXiv:1905.13625] [INSPIRE].

[5] G. Carleo et al., *Machine learning and the physical sciences*, *Rev. Mod. Phys.* **91** (2019) 045002 [arXiv:1903.10563] [INSPIRE].

[6] A. Aurisano et al., *A Convolutional Neural Network Neutrino Event Classifier*, 2016 *JINST* **11** P09001 [arXiv:1604.01444] [INSPIRE].

[7] MicroBooNE collaboration, *Convolutional Neural Networks Applied to Neutrino Events in a Liquid Argon Time Projection Chamber*, 2017 *JINST* **12** P03011 [arXiv:1611.05531] [INSPIRE].

[8] MicroBooNE collaboration, *Deep neural network for pixel-level electromagnetic particle identification in the MicroBooNE liquid argon time projection chamber*, *Phys. Rev. D* **99** (2019) 092001 [arXiv:1808.07269] [INSPIRE].

[9] N. Choma et al., *Graph Neural Networks for IceCube Signal Classification*, in proceedings of the *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Orlando, FL, U.S.A., 17–20 December 2018, pp. 386–391 [arXiv:1809.06166] [INSPIRE].

[10] E. Racah et al., *Revealing Fundamental Physics from the Daya Bay Neutrino Experiment using Deep Neural Networks*, in proceedings of the *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Anaheim, CA, U.S.A., 18–20 December 2016, pp. 892–897 [arXiv:1601.07621] [INSPIRE].

JHEP01(2021)189

[11] EXO collaboration, *Deep Neural Networks for Energy and Position Reconstruction in EXO-200*, 2018 *JINST* **13** P08023 [arXiv:1804.09641] [INSPIRE].

[12] H. Qiao, C. Lu, X. Chen, K. Han, X. Ji and S. Wang, *Signal-background discrimination with convolutional neural networks in the PandaX-III experiment using MC simulation*, *Sci. China Phys. Mech. Astron.* **61** (2018) 101007 [arXiv:1802.03489] [INSPIRE].

[13] P. Ai, D. Wang, G. Huang and X. Sun, *Three-dimensional convolutional neural networks for neutrinoless double-beta decay signal/background discrimination in high-pressure gaseous Time Projection Chamber*, 2018 *JINST* **13** P08015 [arXiv:1803.01482] [INSPIRE].

[14] NEXT collaboration, *Background rejection in NEXT using deep neural networks*, 2017 *JINST* **12** T01004 [arXiv:1609.06202] [INSPIRE].

[15] NEXT collaboration, *Sensitivity of NEXT-100 to Neutrinoless Double Beta Decay*, *JHEP* **05** (2016) 159 [arXiv:1511.09246] [INSPIRE].

[16] D. Nygren, *High-pressure xenon gas electroluminescent TPC for 0-ν $\beta\beta$-decay search*, *Nucl. Instrum. Meth. A* **603** (2009) 337 [INSPIRE].

[17] NEXT collaboration, *Calibration of the NEXT-White detector using ^{83m}Kr decays*, 2018 *JINST* **13** P10014 [arXiv:1804.01780] [INSPIRE].

[18] J. Martín-Albo, *The NEXT experiment for neutrinoless double beta decay searches*, Ph.D. Thesis, University of Valencia, Valencia Spain (2015) [INSPIRE].

[19] GEANT4 collaboration, *GEANT4 — a simulation toolkit*, *Nucl. Instrum. Meth. A* **506** (2003) 250 [INSPIRE].

[20] A. Krizhevsky, I. Sutskever and G.E. Hinton, *Imagenet classification with deep convolutional neural networks*, *Commun. ACM* **60** (2017) 84.

[21] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, *Dropout: A simple way to prevent neural networks from overfitting*, *J. Mach. Learn. Res.* **15** (2014) 1929.

[22] S. Ioffe and C. Szegedy, *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*, arXiv:1502.03167 [INSPIRE].

[23] C. Guo, G. Pleiss, Y. Sun and K.Q. Weinberger, *On calibration of modern neural networks*, arXiv:1706.04599.

[24] K. He, X. Zhang, S. Ren and J. Sun, *Deep Residual Learning for Image Recognition*, arXiv:1512.03385 [INSPIRE].

[25] K. He, X. Zhang, S. Ren and J. Sun, *Identity mappings in deep residual networks*, arXiv:1603.05027.

[26] X. Li, S. Chen, X. Hu and J. Yang, *Understanding the Disharmony Between Dropout and Batch Normalization by Variance Shift*, in proceedings of the *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, U.S.A., 15–20 June 2019, pp. 2677–2685.

[27] J. Deng, W. Dong, R. Socher, L. Li, K. Li and L. Fei-Fei, *ImageNet: A large-scale hierarchical image database*, in proceedings of the *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, U.S.A., 20–25 June 2009, pp. 248–255.

[28] B. Graham and L. van der Maaten, *Submanifold sparse convolutional networks*, arXiv:1706.01307.

JHEP01(2021)189

[29] L. Dominé and K. Terao, *Scalable deep convolutional neural networks for sparse, locally dense liquid argon time projection chamber data*, *Phys. Rev. D* **102** (2020) 012005 [arXiv:1903.05663] [INSPIRE].

[30] C. Shorten and T.M. Khoshgoftaar, *A survey on image data augmentation for deep learning*, *J. Big Data* **6** (2019) 60.

[31] G.J. Székely and M.L. Rizzo, *Testing for equal distributions in high dimension*, *InterStat* **5** (2004) 1.

[32] G. Székely and M.L. Rizzo, *Energy statistics: A class of statistics based on distances*, *J. Stat. Plann. Infer.* **8** (2013) 1249.

[33] R.A. Fisher, *The Design of Experiments*, Oliver and Boyd (1935).

[34] NEXT collaboration, *Sensitivity of a tonne-scale NEXT detector for neutrinoless double beta decay searches*, arXiv:2005.06467 [INSPIRE].

[35] NEXT collaboration, *Initial results of NEXT-DEMO, a large-scale prototype of the NEXT-100 experiment*, 2013 *JINST* **8** P04002 [arXiv:1211.4838] [INSPIRE].

[36] NEXT collaboration, *Operation and first results of the NEXT-DEMO prototype using a silicon photomultiplier tracking array*, 2013 *JINST* **8** P09011 [arXiv:1306.0471] [INSPIRE].

JHEP01(2021)189