ECAP-2019-004

Event Reconstruction for EXO-200 using Deep Learning

Master's Thesis in Physics

Johannes Link
April 26, 2019

Erlangen Centre for Astroparticle Physics
Friedrich-Alexander-Universität
Erlangen-Nürnberg

Supervisors:
PD Dr. Thilo Michel
Prof. Dr. Gisela Anton

ERLANGEN CENTRE FOR ASTROPARTICLE PHYSICS

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

Contents

Introduction

During the past years, the popularity of artificial intelligence in general and Deep Learning, in particular, has grown enormously. From 2014 to 2019, the number of search requests at Google, including "Deep Learning", increased twenty-fold [1]. The number of Deep Learning applications is rising continuously. One reason for this is the growth in computational power of modern hardware. Another reason is the development of better and easier to use machine learning algorithms and frameworks.

The use of Deep Learning also becomes more and more common in the analysis of physical experiments. Because of the very high-dimensional and complex problems of modern particle detectors, classical analysis methods are only possible with very large effort. In addition, a good knowledge of the physical processes inside the detector is crucial. For analyses using Deep Learning, the physical processes in the experiment, are irrelevant. This is both an advantage and disadvantage of Deep Learning. Since the exact processes in the detector are irrelevant the effort for the analysis is much smaller. However, we do not know what the network is doing. Trying to understand the features the network exploits is therefore important to get an idea what the network is doing.

One application of Deep Learning in particle physics is the event reconstruction. In the case of EXO-200, this includes the determination of the position and energy of an interaction. EXO-200 is an experiment searching for the neutrinoless double beta decay. Observing such a decay would answer the question whether the neutrino is a Majorana or Dirac particle. Although the neutrinos where already postulated by Pauli in 1930 [2] and detected by Cowan et al. in the 1950s [3], there are still open questions and therefore subject of recent research. This includes the neutrino mass hierarchy and the masses of the mass eigenstates, which are currently investigated by different experiments [4, 5].

In this thesis, the potential of reconstructing events measured by the EXO-200 detector using Deep Learning methods is investigated. In chapter (1), the theoretical background, including $\beta\beta0\nu$ physics, Deep Learning, and the EXO-200 experiment, is introduced. The Convolutional Neural Network used and the data sets used within this thesis are the subjects of (2). Chapter (3) discusses the reconstruction of Monte Carlo simulated events and chapter (4) discusses the performance of the network on real data.

CHAPTER 1

Theoretical Background

Contents

In this chapter, the theoretical background of this thesis is outlined. This includes the physical background i.e. the neutrino accompanied and the neutrinoless double beta decay. It is followed by the setup of the EXO-200 detector with its detection principle and standard event reconstruction. Furthermore, the concepts of Deep Learning including backpropagation and convolutional neural networks are explained.

For a more detailed introduction to the different topics see the sources given or special literature. If you are already familiar with the covered topics feel free to skip this chapter.

1.1 Physical Background

In this section, the important physical fundamentals of this thesis are presented. Based on the double beta decay, the hypothetical neutrinoless double beta decay is explained and the implications of an existence of this decay are shown.

1.1.1 Double Beta Decay

The beta decay is a radioactive decay of an atomic nucleus ${}^{A}_{Z}X$ with mass number A and atomic number Z. In this reaction, a neutron of the nucleus decays into a proton

$$ {}^{A}_{Z}X \rightarrow {}^{A}_{Z+1}Y + e^{-} + \bar{\nu}_e \, . \tag{1.1} $$

The new nucleus Y has the same mass number A but an increased atomic number $Z+1$. Furthermore, one electron and one electron antineutrino are emitted. The binding energy of the daughter nucleus Y is lower than the one of the mother nucleus. Due to that, energy is released during the beta decay.

For some particles, this single beta decay is energetically forbidden. This is the case for nuclei with larger binding energy of the hypothetical daughter nucleus. Several of these nuclei react via a second-order decay called the double beta decay ($\beta\beta 2\nu$), where two electrons and two electron antineutrinos are emitted and the atomic number of the daughter nucleus increases by two

$$ {}^{A}_{Z}X \rightarrow {}^{A}_{Z+2}Y + 2e^{-} + 2\bar{\nu}_e \, . \tag{1.2} $$

As the lepton number is preserved during this decay, it is permitted within the Standard Model of particle physics. Figure 1.1a shows the Feynman diagram of the double beta decay. In 1935, the double beta decay was first described by M.Goeppert-Mayer [6] and in 1950 it was measured for the first time by M. G. Inghram and J. H. Reynolds [7] in the nucleus ^{130}Te via geochemical methods. Modern measurements of the half-life are of the order of $T_{1/2} \approx 10^{18}\,\mathrm{yr} - 10^{21}\,\mathrm{yr}$ [8]. This makes the double beta decay the rarest decay detected so far [8]. The sum of the kinetic electron energy emitted during this process follows the broad example spectrum shown in Figure 1.2.

1.1.2 Neutrinoless Double Beta Decay

In contrast to the neutrino accompanied double beta decay, the neutrinoless double beta decay ($\beta\beta 0\nu$) is a hypothetical decay that has not been observed yet. It was first postulated by W. H. Furry in 1939 [10]. The Feynman diagram for the neutrinoless double beta decay is shown in Figure 1.1b. Simplified the electron

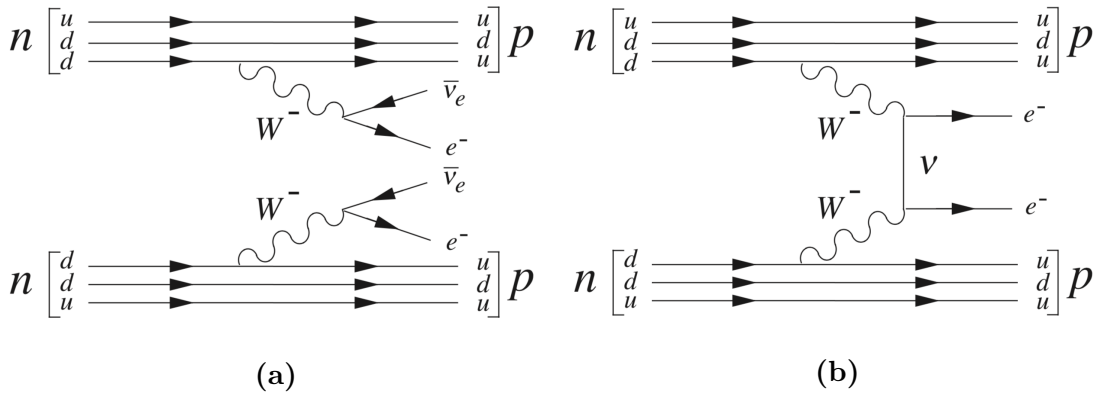

Figure 1.1: Feynman diagram of **(a)** the double beta decay, where two neutrons decay into two protons. During this process, two electron antineutrinos and two electrons are emitted. Feynman diagram **(b)** shows the neutrinoless double beta decay where in contrast only two electrons are emitted. Reprinted from [9].

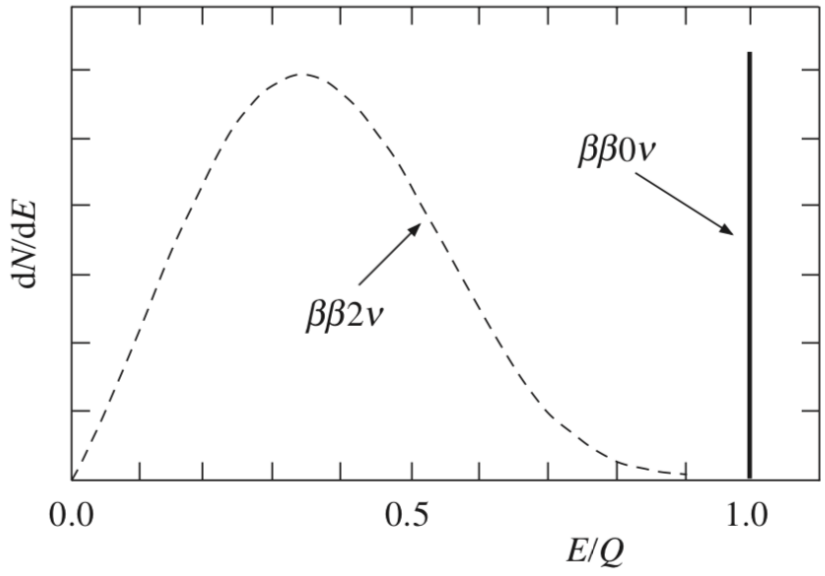

Figure 1.2: Energy spectra for the two electrons of the double beta decay and the neutrinoless double beta decay as a function of E/Q, where E is the combined energy of the electrons and Q is the total energy released. Reprinted from [9].

antineutrino emitted by the double beta decay at one vertex, is absorbed as an electron neutrino at the second vertex. This leads to the following reaction equation

$$ {}^{A}_{Z}X \rightarrow {}^{A}_{Z+2}Y + 2e^{-} \ . \tag{1.3} $$

Physically correct the process is much more complex [11], but Schechter and Valle showed in 1982 that the conclusion $\nu = \bar{\nu}$ is independent of the exact mechanism [12]. This reaction is only possible if some requirements are fulfilled. First of all, the neutrino must be a Majorana particle i. e. its own antiparticle. Another point is that the neutrino must have mass. This enables a helicity flip from a right-handed antineutrino to a left-handed antineutrino. Through the observation of neutrino oscillations, at least two neutrino mass eigenstates must have a nonzero rest mass. Since only two electrons are emitted at the neutrinoless double beta decay the expectation is that these electrons have a sharp sum energy of $E = Q$. To detect the hypothetical decay one is looking for this so-called Q value. Figure 1.2 shows the different electron energy sum spectra for the double beta decay and the neutrinoless double beta decay.

For an experiment observing no neutrinoless double beta decay, only a lower limit for the half-life of the decay can be given. The effective neutrino Majorana mass $m_{\beta\beta}$ is defined as [8]

$$ m_{\beta\beta} = \left| \sum_{i=1}^{3} = m_i \cdot \left(U_{e,i}^{\text{Majorana}} \right)^2 \right| \ , \tag{1.4} $$

where m_i is the mass of the mass eigenstates ν_i and $U_{e,i}^{\text{Majorana}}$ the elements of the Majorana PMNS matrix for the electron flavor. The half-life and the effective neutrino Majorana mass $m_{\beta\beta}$ are related via [13]

$$ \frac{1}{T_{1/2}^{\beta\beta0\nu}} = G_{0\nu}(Q_{\beta\beta}, Z) \cdot |M_{0\nu}|^2 \cdot m_{\beta\beta}^2 \ , \tag{1.5} $$

where $M_{0\nu}$ is the nuclear matrix element and $G_{0\nu}(Q_{\beta\beta}, Z)$ is the phase space factor of the neutrinoless double beta decay. So, a lower limit on the half-life corresponds to an upper limit on the effective neutrino Majorana mass.

If an experiment has no background close to the Q value, the upper limit on the effective neutrino Majorana mass $m_{\beta\beta}$ after an exposure $M_{\beta\beta} \cdot t$ is [14]

$$ m_{\beta\beta} \propto (M_{\beta\beta} \cdot t)^{-\frac{1}{2}} \ , \tag{1.6} $$

where $M_{\beta\beta}$ is the mass of the $\beta\beta$ nucleus of interest and t the measurement period. One has to underline that the lower limit on the half-life scales with the square-root of the exposure.

For an experiment with background in the region of interest the relation to the exposure changes to [14]

$$m_{\beta\beta} \propto (M_{\beta\beta} \cdot t)^{-\frac{1}{4}} \ . \tag{1.7}$$

Because of this relation, it is very important to understand and minimize the background of an experiment searching for the neutrinoless double beta decay.

As the neutrinoless double beta decay was not observed yet, one can only give a lower limit for the half-life. The current lower limits exceed $10^{25} - 10^{26}$ yr depending on the nucleus [15].

1.1.3 Physical Implications

The detection of the neutrinoless double beta decay implies to physics beyond the Standard Model and answers some open questions in neutrino physics. A first consequence of a detection is that the neutrino is a Majorana particle, i. e. its own antiparticle. This fact could lead to a better understanding of the matter-antimatter asymmetry problem.

Secondly, that the neutrinos have mass, which is already confirmed through the detection of neutrino oscillations. Apart from this fact the exact masses and the ordering of the mass eigenstates is still not known. This is due to the fact that through neutrino oscillations only the absolute value of the squared mass differences $|\Delta m^2_{ij}|$ between the mass eigenstates ν_i and ν_j can be measured [16]. Additionally, it is known that the mass eigenstate ν_1 has a smaller mass than the mass eigenstate ν_2. Figure 1.3 shows the two possible neutrino mass hierarchies which are the so-called normal hierarchy and the inverted hierarchy. The difference between these hierarchies is that in the inverted hierarchy the mass eigenstate ν_3 has a smaller mass than the mass eigenstates ν_1 and ν_2.

Through a measurement of the half-life of the neutrinoless double beta decay, it is possible to constrain the absolute neutrino mass and the mass hierarchy.

1.2 The EXO-200 Experiment

In this section, the setup of the EXO-200 experiment and its detection principle is explained. Furthermore, the main steps of the event reconstruction are explained.

The Enriched Xenon Observatory (EXO) is a particle experiment searching for the neutrinoless double beta decay. It is located at the Waste Isolation Pilot Plant, WIPP, in New Mexico, USA, which is a radioactive waste repository [18]. At WIPP, the detector is situated at a depth of 655 m which corresponds to 1858 meters water equivalent [19]. The reason the experiment was conducted in a location with much overburden is to have less atmospheric background. The majority of the cosmic rays are absorbed by the rock over the salt mine.

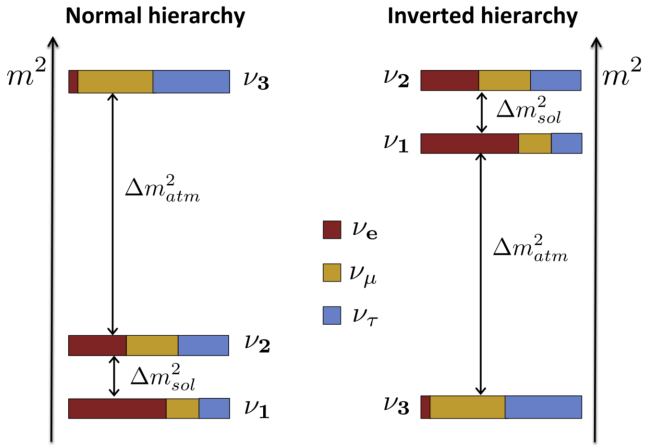

Figure 1.3: Neutrino mass eigenstates ν_1, ν_2, and ν_3 arranged in the normal (left) and inverted (right) neutrino mass hierarchy. The composition of the mass eigenstates through the flavour eigenstates ν_e, ν_μ, and ν_τ is represented by the ratio of the colored bars. Reprinted from [17].

EXO-200 was conducted in two phases (2011 – 2014, 2016 – 2019) and is a prototype detector to test the technology of single-phase time projection chambers (TPCs) for a future, larger detector (nEXO). The idea is to use the xenon as both source and detection material. The xenon used is enriched in ^{136}Xe (80.6%) and the active detector volume is about 110 kg of liquid xenon [20].

1.2.1 Setup

The EXO-200 detector consists of a single-phase double-sided TPC filled with liquid xenon. Both sides of the TPC are of cylindrical shape with a height of 40 cm and a diameter of 44 cm and share the same cathode.

Figure 1.4 sketches the TPC. The vessel consists mainly of copper and bronze, which was degreased and etched to not contaminate the liquid xenon in the vessel [20].

The inner shell surface of the cylinder is covered with teflon to reflect scintillation light. To detect the scintillation light large area avalanche photodiodes (LAAPDs) are mounted at the two bases of the cylinder. Between the two sides of the TPCs there is a cathode biased at −8 kV for phase I and at −12 kV for phase II. The so-called field shaping rings located at the shell surface of the cylinder provide a homogeneous electrical field in the active volume. At both end caps of the TPC, a shielding (V) and a charge collection (U) wire plane are located respectively. The V-wire planes are at a voltage of −780 V [21]. The shielding, or induction, wire planes are mounted in front of the U-wires. Each wire plane consists of 38 wire

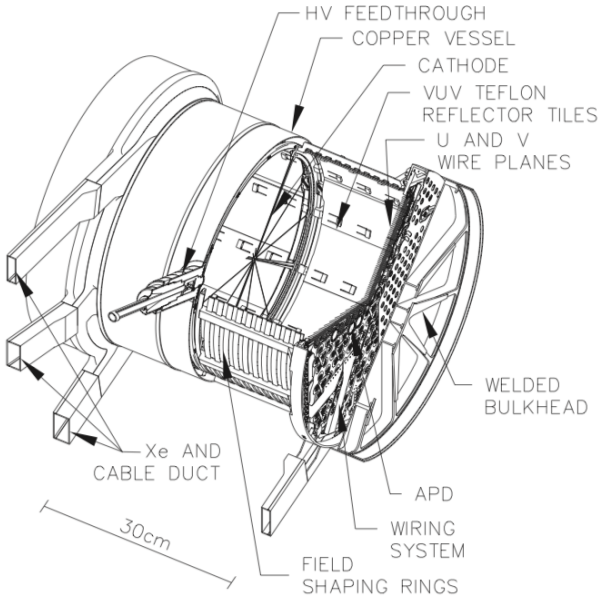

Figure 1.4: Sketch of the EXO-200 detector with a cutaway view to see the inner parts. All important elements are labelled. Reprinted from [20].

triplets. The wire planes of each TPC form a hexagon together which is illustrated in Figure (1.5).

The TPC is surrounded by a double-walled cryostat to keep the liquid xenon at a temperature of 167 K and a pressure of 147 kPa. The cryostat again is housed inside a 25 cm lead shielding and surrounded by cosmic-ray veto panels. These veto panels are made of plastic scintillators and are used to identify cosmic and atmospheric background [20].

1.2.2 Detection Principle

An event taking place in the detector leads to excited Xe atoms, Xe^*, or electron-ion pairs, $Xe^+ + e^-$. Figure 1.6 shows possible excitation processes.

Through the energy deposit in the detector, free electrons and scintillation light is created. The event energy deposited in the scintillation and ionization channel is anti-correlated. Combining these two channels leads to a better energy resolution [23].

In Figure 1.7, the process of the electron drift is shown schematically. The created free electrons drift, due to the electric field in the TPC, towards the anode. Because of the difference of potential between the U- and V-wires, the V-wires are transparent for the electrons. Thus, the electrons are not collected on the

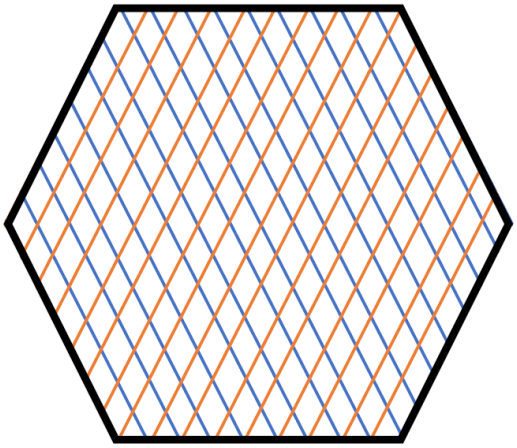

Figure 1.5: Scheme of the end of the TPC with the wire planes. The U-wires **(orange)** and the V-wires **(blue)** are crossed by 60° and form a hexagon. Each line illustrates a wire triplet. The Number of wires in this sketch does not match the real number of wires.

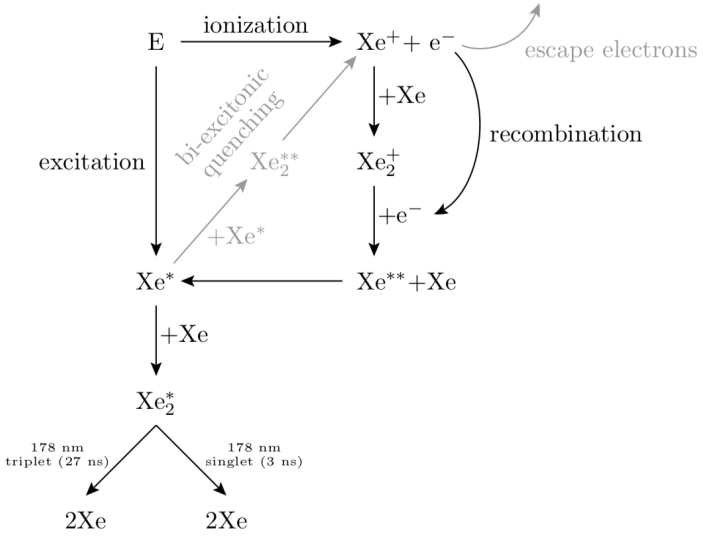

Figure 1.6: Scheme of the process leading from an energy deposit in liquid xenon to scintillation light and ion electron pairs. Reprinted from [22].

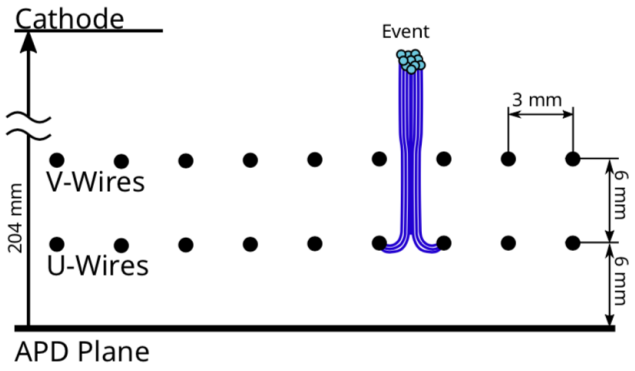

Figure 1.7: Scheme of an event taking place in the detector. The drifting electrons pass the V-wire plane and lead to an induction signal. Afterwards the electrons are collected at the U-wires. Reprinted from [24].

V-wires but only produce an induction signal in these wires. The electrons are then collected by the U-wire plane. The wires in both wire planes are aligned at an angle of 60° to each other. Due to this, one can measure both a U- and a V-coordinate of an event. These coordinates are unique and can be converted to Cartesian coordinates.

The scintillation light, created by an energy deposit in the detector is reflected by the teflon tiles at the shell surface and measured by the LAAPDs at the respective end of the TPC.

Using the known drift velocity, the time difference between the light signal and the charge signal is used to determine the Z-coordinate of an event. The electron lifetime in the liquid xenon depends on the purity of the xenon in the TPC. Due to that, the current purity must be known to determine the deposited energy.

1.2.3 Classical Event Reconstruction

The event reconstruction can be divided into three parts. Firstly, the signal finding, followed by the signal parameter estimation, and finally the clustering of found signals to events. A short overview of the different parts is given in this section. For a more detailed explanation of the event reconstruction see [25].

To find a signal in the constant data stream a matched filter is applied to the APDs, U-, and V-wires. A signal is found if the response to the matched filter exceeds a certain threshold. This threshold is not at a fixed value but depends on the root mean square of the noise in order to be sensitive to noise variations. After the matched filter stage an “unshaping” algorithm is applied, which disentangles possible multiple signals which are temporally close to each other.

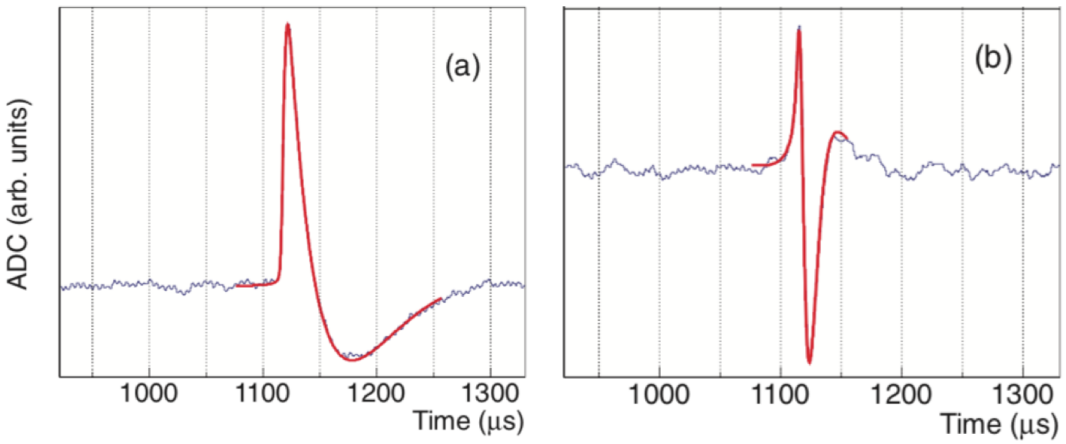

Figure 1.8: Example waveforms of a signal on a U-wire **(a)** and a V-wire **(b)**. The red curve shows the signal template fitted to data. The peak of the V-wire signal is earlier in time than the peak in the U-wire signal. Reprinted from [25].

For the U, V, and APD channels appropriate signal templates are fitted to the data to obtain the signal parameters. Figure 1.8 shows an example signal on the U- and V-wires with a fit of the respective signal template.

The last part of the event reconstruction is the event clustering. Firstly, the signals from the same event of the channels of the same type are bundled together. By combining the U- and V-wire information, a 2D position of a charge cluster is determined. After that, the same event bundles of the U, V, and APD channels are clustered together to get the full 3D position of the charge cluster, in which one scintillation cluster can contain many charge clusters.

The EXO-200 standard analysis distinguishes between single-site (SS) and multi-site (MS) events. Single-site events are defined as events with one charge cluster, where only one or two U-wires collect charge. Multi-site events are all events that are not single-site events. This classification is physically motivated by the classification of background and signal as MS events are background-like and SS events are signal-like.

1.3 Deep Learning

In this section, the basic principles of Deep Learning are explained. Deep Learning methods can be divided into three different types. One type is supervised learning, in which the networks receive the correctly assigned output label for each input during training. Another one is unsupervised learning, where the network does not get the correctly assigned output label but e. g. tries to find patterns in data

to cluster them. The third type is reinforcement learning where an agent reacts to its environment and is rewarded depending on its action. This can be used to control robots or to solve games. As in this thesis only supervised Deep learning techniques are applied, this section focuses particularly on this type.

1.3.1 Basics of Deep Learning

A Deep Neural Network consists of many neurons usually arranged in layers, whereby each neuron applies the following operation to its inputs x_i

$$f(x_1, x_2, ..., x_N) = \varphi\left(\sum_{i=1}^{N} x_i w_i + \Theta\right) , \tag{1.8}$$

where w_i is the corresponding weight to the input x_i from the previous layer, Θ is the bias, and φ is the activation function. The latter introduces a nonlinearity to the network. Currently, the most popular activation function is the rectified linear unit (ReLu) which is defined as

$$\varphi(x) = \max(0, x) . \tag{1.9}$$

Figure 1.9 shows a simple neural network with four input neurons and two output neurons. All layers between the input and output layer are called hidden layers. If every neuron of one layer is connected to every neuron of the following layer, the layer is called a fully connected layer. A feedforward network is a network, where the connections do not form a cycle, but point in the same direction. The information in a feedforward network always moves in one direction.

It is proven that a feedforward network with only one hidden layer and a sufficient number of neurons in this hidden layer can approximate any continuous function on a compact subset of $\mathbb{R}^n$ [27]. But this is not feasible in practice, as an enormous number of neurons would be needed, which can be reduced by arranging the neurons in many layers.

1.3.2 Backpropagation

By minimizing an error function and adapting the weights of the neurons, networks learn according to the value of the loss function. For regression tasks, the mean absolute error or the mean squared error are typically used as an error function. The minimization of the loss function is done using the so-called backpropagation which is shortly introduced in this section and is adapted from [28].

Using the chain rule, the gradient of a loss function E with respect to a certain weight w_{ji} can be written as

$$\frac{\partial E}{\partial w_{ji}} = \frac{\partial E}{\partial a_j} \frac{\partial a_j}{\partial w_{ji}} = \delta_j z_i , \tag{1.10}$$

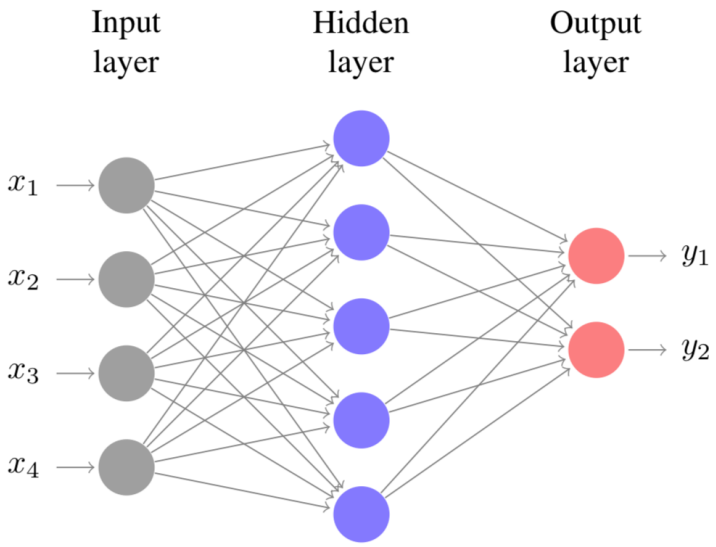

Figure 1.9: Scheme of a fully connected feed forward network with four input neurons and two output neurons. The network has one hidden layer consisting of five neurons. Each of the connecting lines illustrates a weight w_i. Reprinted from [26].

where $z_i = \varphi(a_i)$ is the activation of neuron i, with a_i the weighted sum of the inputs as in equation (1.8), and δ_j is the "error". From equation 1.10 it can be seen that the derivative of the error function E can be calculated by multiplying the value of δ at the output end of the connection by the value of z on the input end.

With $\varphi' = \frac{\partial \varphi}{\partial a}$, the "error" of the output neurons can be calculated as

$$\delta_k \equiv \frac{\partial E}{\partial a_k} = \varphi'(a_k) \frac{\partial E}{\partial y_k} \,. \tag{1.11}$$

The "error" of the hidden neurons can be calculated using the chain rule as well and by summing over all neurons k to which neuron j sends its output

$$\delta_j = \varphi'(a_j) \sum_k w_{kj} \delta_k \,. \tag{1.12}$$

With this back-propagation formula, the derivative of the loss function with respect to the weights is calculated during training. These weights are then updated using the gradient descent technique

$$w_{ji}^{\text{new}} = w_{ji}^{\text{old}} + \Delta w_{ji} \,. \tag{1.13}$$

In practice batch training is very common. This means that the weights are not updated after each input but after a batch of inputs. So, the change of the weight

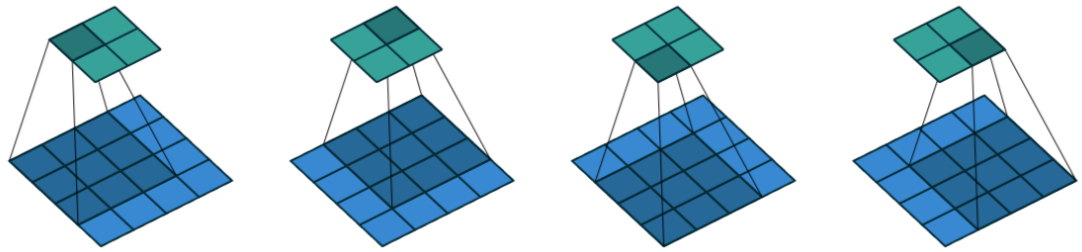

Figure 1.10: Convolution of a 4x4 input image with a 3x3 kernel resulting in a 2x2 output. The 3x3 kernel consists of nine weights. No padding and stride 1 is applied. Reprinted from [29].

w_{ji} can be written as:

$$\Delta w_{ji} = -\eta \sum_{n} \delta_j^n x_i^n \ , \tag{1.14}$$

where the sum contains all inputs x^n of a batch and η is the learning rate, which is used as a hyper parameter during training.

1.3.3 Convolutional Neural Networks

Convolutional neural networks are a class of neural networks very common for image recognition tasks.

A Convolutional neural network typically consists of convolutional layers followed by fully connected layers. In a convolutional layer, the input image is convolved with different kernels. This descriptively corresponds to the kernels being shifted over the input image and a weighted sum of the input image in the receptive field of the kernel and its weights is computed. Each kernel has different weights and for each kernel another output, the so-called feature map is obtained. Figure 1.10 depicts the procedure of a convolution. The different kernels are looking for different features in the input image. The first convolutional layers usually are sensitive to simple features like edges and the deeper layers to more abstract and complex features.

The stride of a convolution is the number of pixels the kernel is shifted during the convolution. To conserve the size of the input layer during convolution padding can be applied. Zeros are added around the input image so that the output shape is equal to the previous input shape.

Between two convolutional layers, a max-pooling layer is used often. This layer outputs the maximum response value over a receptive field. This is done to reduce the dimension of the feature maps and to have fewer weights, which reduces computational cost.

The advantage of a convolutional layer in comparison to a fully connected layer is that the number of weights is much smaller. Another advantage is that the network can exploit translational invariance which means that it can recognize patterns e. g. in images independent of the location in the image. This is why convolutional neural networks are the standard architecture for image recognition.

CHAPTER 2

Deep Learning for EXO-200

Contents

In this chapter, the data production using a Monte Carlo simulation and the data set properties are outlined. Furthermore, the architecture and the training of the neural networks used to reconstruct the events in the EXO-200 detector are introduced.

2.1 Monte Carlo Data

The data used to train the neural network is generated in a GEANT4 [30] based Monte Carlo simulation of the EXO-200 detector. Using the simulation, labeled events are generated, which is necessary for supervised training of a deep neural network (DNN). The functionality of the Monte Carlo simulation is described in the next subsection. The simulation generates raw signals of the wires in the U- and V-wire planes. These are then preprocessed and finally used as an input for the network. In the preprocessing step, the individual gain factors of the detection wires are corrected. This is done by dividing the simulated signals of each channel according to the channel-specific measured gain factor.

Figure (2.1) shows an example of the simulated measurement of the U- and V-wire planes. The signal, in the two left wire planes of the figure, clearly stands out from the noise. The transparent part in the figure is cropped, as this part only

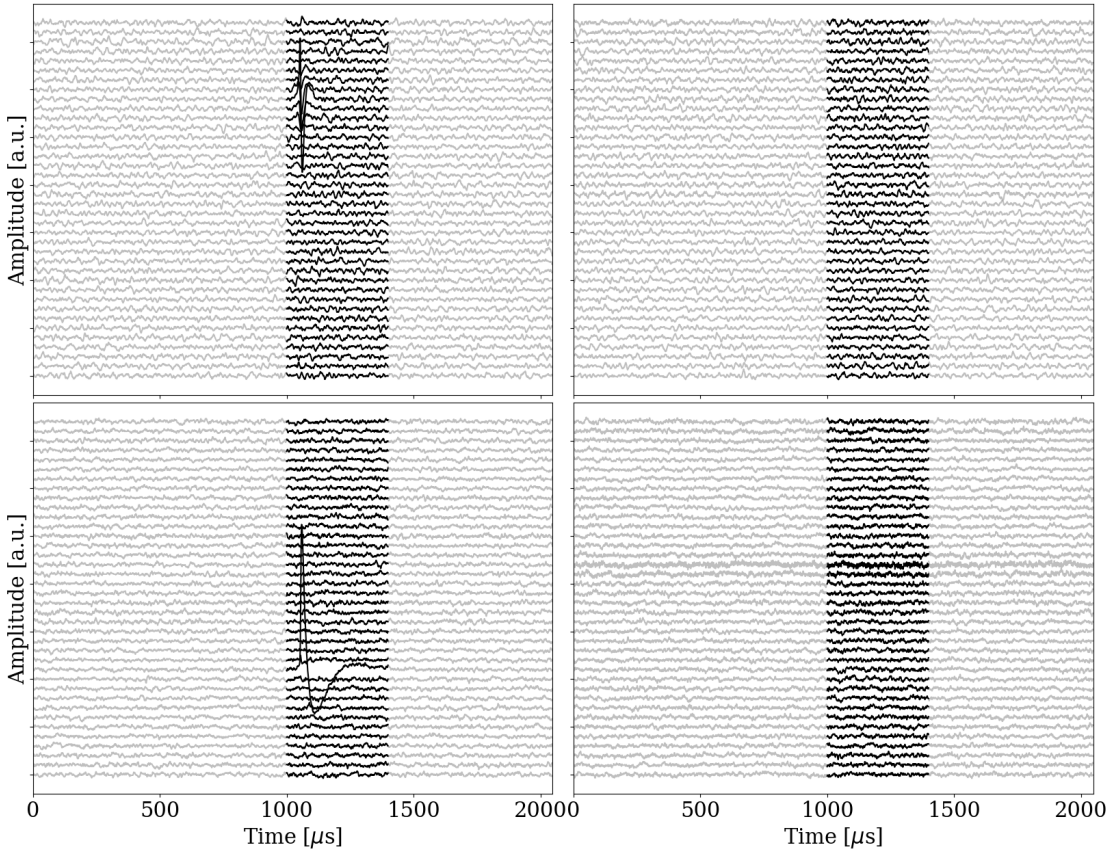

Figure 2.1: Signal of the V-wires (top) and the U-wires (bottom) of the two sides of the TPC (right/left). The signals are shifted by an offset to visualize all wires. Before the image is used as an input for the network, the signals are gain-corrected and the transparent area is cropped as it only contains noise.

contains noise and no information about the event. This can be done because in the Monte Carlo simulation the signal is always in the same range after the trigger. Hence, fixed values (1000 µs − 1400 µs) for the cropping can be used.

2.1.1 Monte Carlo Simulation

The simulation procedure for EXO-200 is run in two separated steps. In the first step, the geometry of the detector is modeled in GEANT4. GEANT4 is a toolkit to simulate particle physics interactions with matter [30]. Using this toolkit, the energy depositions of physical interactions in the liquid xenon are simulated and sampled in pixelated charge deposits (PCDs) of cubic voxels with an edge length of 0.2 µm.

In the second part of the simulation, the electronic signal from the PCDs is generated. The deposited energy leads to free electrons and scintillation light. The free electrons are drifted in the electric field within the detector. During the timestep Δt, a PCD, combining these free electrons, moves $\Delta\vec{x}$ in the opposite direction of the electric field $\vec{E}$

$$\Delta\vec{x} = -v_{\mathrm{D}} \cdot \Delta t \cdot \frac{\vec{E}}{|\vec{E}|} \,. \tag{2.1}$$

The drift velocity v_{D} is constant and determined from measurements. The diffusion of the electrons, during the drift, is also included in the simulation. Additionally, the electron lifetime in the liquid xenon can be adjusted in the simulation, to simulate different xenon purities.

Using the Shockley-Ramo theorem, the induction on the detection wires from the drifting electrons is determined [25, 32, 31]

$$i = E_v \cdot q \cdot |\vec{v}| \,, \tag{2.2}$$

where i is the current induced, E_v is the electric field in the direction of $\vec{v}$, q is the charge of the moving particle, and $\vec{v}$ its velocity. After this step, the transfer function of the electronics is applied to the signal. Finally, real noise, recorded with the detector, is added [33].

2.1.2 Training Data

For training the DNN, a data set of 208,000 events, produced with the Monte Carlo simulation, is used. It is simulated with a uniform distribution in energy and in position in the detector. After considering energy dependent cross sections and applying some event selection cuts, the uniformity is slightly degraded. The true position label is calculated as the energy-weighted mean position of the PCDs. The uniform distribution is chosen, that the network learns to reconstruct all events without introducing a bias, independent of the energy and position of the event. The events are simulated for the measuring phase I of EXO-200 which has a different electric field as phase II and only SS events are used. Figure (2.2) shows the correlation matrix of the later target parameters U, V, Z, and E of the training data. On the diagonal plots, the distributions of the four parameters are shown. Because of the hexagonal shape of the wire planes, the U- and V-coordinate are correlated and not uniformly distributed, even if the events have a uniform distribution in the detector volume. The off-diagonal elements contain two-dimensional distributions with every combination of the parameters. Except for the correlation between the U- and V-coordinate, the parameters are not correlated.

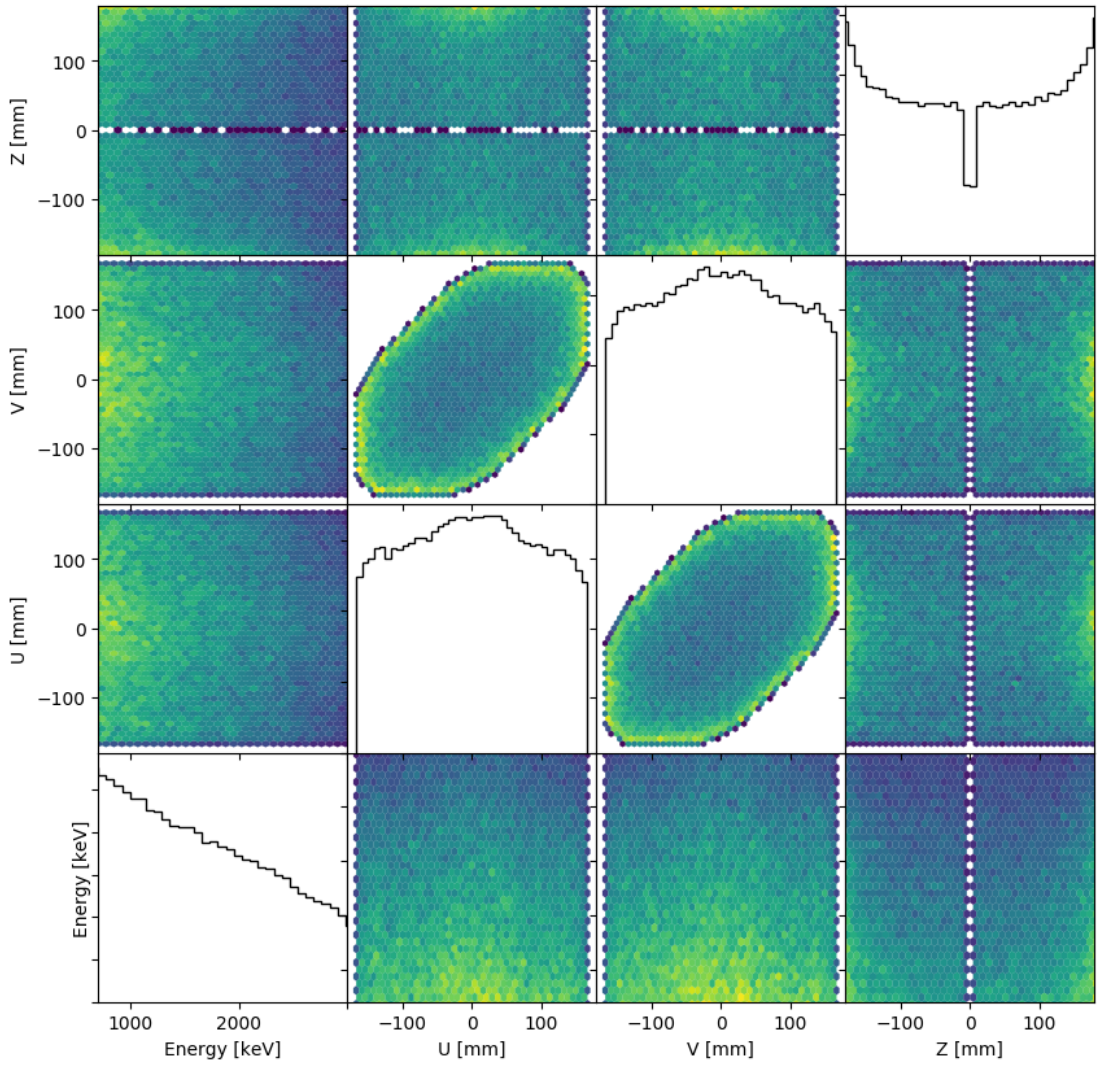

Figure 2.2: Correlation matrix of the training data set. The diagonal elements show distributions of the target parameters, whereas the off-diagonal elements show the parameters as a function of another parameter. Apart from the correlation of the U- and V-coordinate, due to the hexagonal shape, the remaining parameters are uncorrelated.

Source location	nominal position (x, y, z)[mm]	true position (x, y, z)[mm]
S2	$(0, 0, -29.5)$	$(-2.54, 0.23, -29.2)$
S5	$(25.5, 0, 0)$	$(25.5, 0.39, -3.0)$
S8	$(0, 0, 29.5)$	$(3.5, 0, 29.5)$

Table 2.1: Coordinates of different calibration source positions. The true positions differ from their nominal values. Values from [18].

The training data is simulated with an infinite electron lifetime, i. e. no impurities within the liquid xenon. In reality, the electron lifetime is finite and depends on the xenon purity. An infinite electron lifetime was chosen for the training data because the purity of the xenon is not constant over time. When applying the network to data with finite electron lifetime, the reconstructed energy has to be corrected.

The training set consists of gamma events. Furthermore, the following selection cuts are applied to the training data. Firstly, the total energy deposited by an event in the liquid xenon must be between 700 keV and 3000 keV. Secondly, the distance between the active PCDs of an event must be smaller than 5 mm. Active means in this context that the energy deposited in the liquid xenon is also deposited on one of the detection channels and not lost in some inactive detector component.

2.1.3 Test Data

To validate the performance of the network, Monte Carlo simulated events of $\beta\beta0\nu$ decays and of calibration sources at different positions are used. As a calibration source ^{228}Th with its prominent peak of ^{208}Tl at $E = 2614.5$ keV near the Q-value is used. Table (2.1) and Figure (2.3) show the coordinates of the different positions. The real positions of the calibration sources differ slightly from their nominal values. The MC simulation uses the true positions of the calibration sources. Unlike the training data, the calibration data sets are not uniformly distributed, but have real physical spectra in energy and position. The $\beta\beta0\nu$ data set has a uniform distribution in space but a sharp peak at $E = 2458$ keV.

Apart from the simulated Monte Carlo Data, real calibration data is used to validate the network's performance. Both, simulated and real events, are from measuring phase I of EXO-200 and are only SS events.

Since in reality, the secondary electrons in the xenon have a finite lifetime, the simulated calibration source data was also simulated with a finite electron lifetime of $\tau = 4500$ µs which matches the average lifetime measured in phase I. This is done to have a better comparison of the networks' performance on real and simulated data.

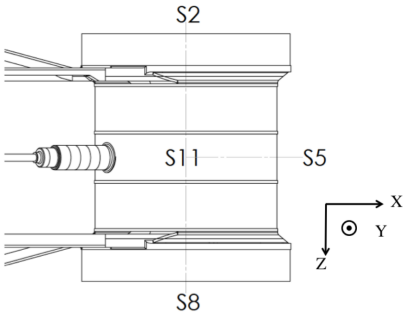

Figure 2.3: Different calibration source positions used for validation of the network. Reprinted from [18].

Because the electron lifetime affects the energy reconstruction, the predicted energy E_{DNN} must be corrected using the formula [25]

$$E_{\mathrm{DNN}}^{\mathrm{corrected}} = E_{\mathrm{DNN}} \cdot \exp\left(\frac{t_{\mathrm{drift}}}{\tau}\right) , \tag{2.3}$$

where $E_{\mathrm{DNN}}^{\mathrm{corrected}}$ is the purity corrected energy, t_{drift} is the drift time of the electrons and τ is the lifetime of the electrons in the liquid xenon. The drift time t_{drift} depends on the z coordinate of the event and is calculated by

$$t_{\mathrm{drift}} = \frac{d_{\mathrm{C}\to\mathrm{APD}} - d_{\mathrm{APD}\to\mathrm{U}} - d_{\mathrm{U}\to\mathrm{V}} - |z_{\mathrm{pred}}|}{v_{\mathrm{drift}}} + t_{\mathrm{collection}} , \tag{2.4}$$

where $d_{\mathrm{C}\to\mathrm{APD}} \approx 204.4\,\mathrm{mm}$ is the distance between the cathode and the APD plane, $d_{\mathrm{APD}\to\mathrm{U}} = 6.0\,\mathrm{mm}$ the distance between the APDs and the U-wire plane, $d_{\mathrm{U}\to\mathrm{V}} = 6.0\,\mathrm{mm}$ the distance between the U- and V-wire plane, z_{pred} the z-position predicted by the reconstruction method, $v_{\mathrm{drift}} = 0.00171\,\frac{\mathrm{mm}}{\mathrm{ns}}$ the drift velocity of the secondary electrons and $t_{\mathrm{collection}} = 2940.0\,\mathrm{ns}$ the collection time. The collection time is the drift time of the secondary electrons from the V- to the U-wire plane.

2.2 Network Architecture

Convolutional neural networks are state-of-the-art networks to analyze images. The measurement data from the induction and collection wires of the EXO-200 detector can be interpreted as a grayscale image. Because of this, a convolutional neural network is used to reconstruct the events in the EXO-200 detector. Figure 2.4 shows a sketch of the setup of the network. All layers in the network use ReLu as activation function and are implemented in Keras [34] using Tensorflow [35] backend.

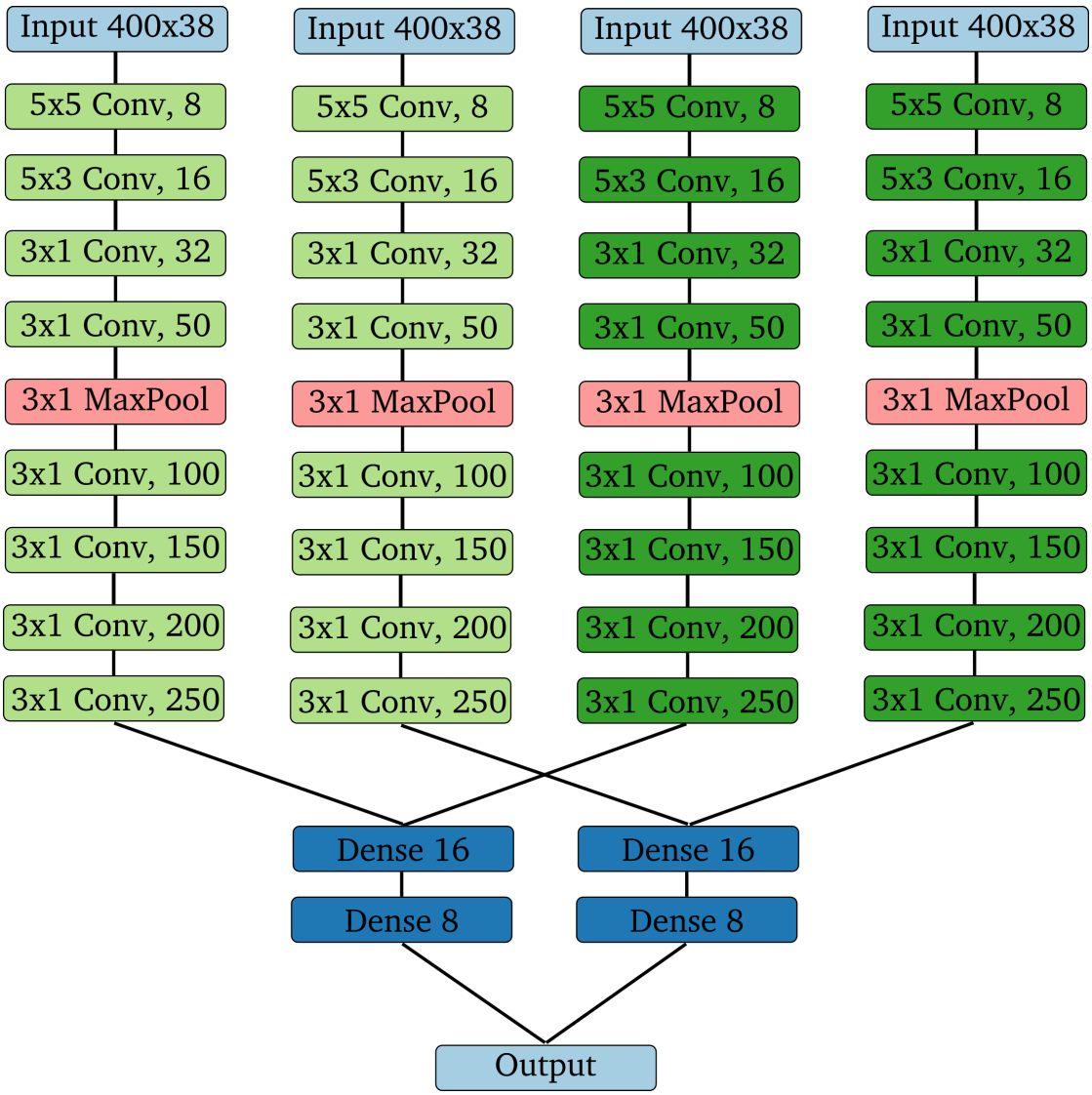

Figure 2.4: Sketch of the convolutional neural network. The four input images are the U- and V-wire signals of both sides of the TPC. The Network branches, illustrated in the same color, share their weights.

As an input, the cropped and gain-corrected waveforms of both V-wire and both U-wire planes are used. Each consists of 38 wire channels and is cropped to 400 samples. Eight convolutional layers are applied to each of the input images. Whereby the kernels applied to the U-wires share their weights and the kernels applied to the V-wires share their weights. This is done to have less training parameters, which corresponds to less computational cost. It is physically motivated, as both sides of the TPC have the same geometry. Network branches with shared weights are represented by the same colors in the sketch of the network. After four convolutional layers, a max pooling layer is applied to the U- and V-wire branch. The reason for this layer is also to save computational cost. After the convolutional layers, the feature maps of the wire planes of the respective side of the TPC are separately concatenated and flattened. Due to this, for each side of the TPC, the stack of two-dimensional feature maps becomes a one-dimensional array. To each of the one-dimensional arrays, two fully connected layers are applied. The neurons of the fully connected layers for both sides of the TPC share their weights. The outputs of the last dense layer are concatenated and used as an input for the output neuron.

A DNN with this architecture is used for each of the four parameters, the U-, V-, and Z-coordinate as well as the energy. The network is trained for each parameter individually to get the best results possible with this architecture.

2.3 Network Training

The training of the networks is done on a graphics processing unit (GPU) at the Regionales Rechenzentrum Erlangen (RRZE) using the Adam optimizer [36].

During training, the training data introduced in subsection (2.1.2) is split into a training and a validation set. The training set consists of 90% of the data and the validation set of the remaining 10%. The network does not see the validation data during the training. This is done to check with an independent data set if the network is overfitting the training data. A network is overfitting the training data if it is too adapted to the training data and therefore becomes worse in generalizing to new data.

The mean squared error (commonly known as L2-loss) is used as a loss function. As an example Figure (2.5) shows the loss function as a function of the training time for the network reconstructing the V coordinate. The value of the loss function for the training data set (blue) and validation data set (green) is, apart from statistical fluctuations, identical, which implies, that the network is not overfitting the training data.

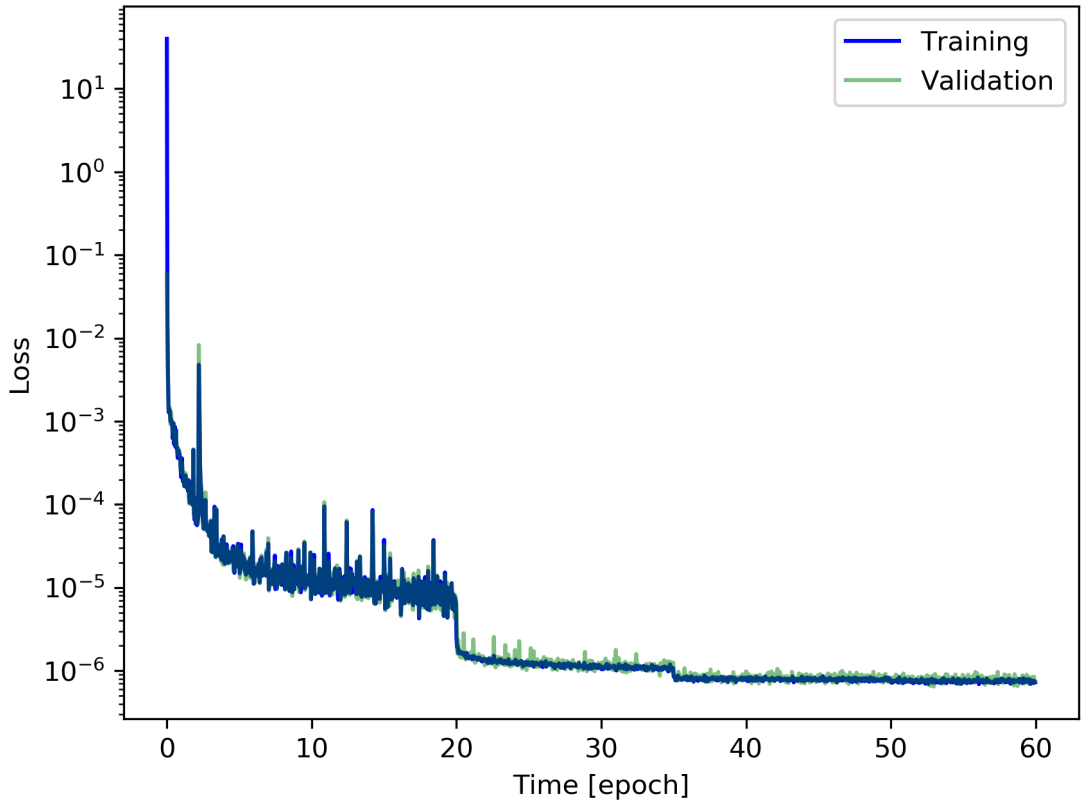

Figure 2.5: Loss plotted as a function of time in epochs during training on the V-coordinate both for training data **(blue)** and validation data **(green)**. The learning rate has been decreased by a factor of ten after epoch 20, 35 and 50. This leads to the jumps in the loss.

Starting with a learning rate of 10^{-3}, the learning rate is decreased by a factor of ten after epoch 20, 35 and 50. This results in jumps in the loss function. The learning rate is reduced in order to reach the minimum of the parameter space.

For too large values of the learning rate, the network jumps over the minima of the parameter space but does not reach them. For too small learning rates the training takes very long and gets stuck in local minima.

CHAPTER 3

Reconstruction of Simulated Events

Contents

In this chapter, the performance of the trained convolutional neural networks in reconstructing the energy and position of simulated single-site (SS) events is investigated and compared to the classical analysis of EXO-200. For this purpose approximately 60,000 $\beta\beta0\nu$ SS events, uniformly distributed within the detector, are simulated using the EXO-200 Monte Carlo simulation. The uniform event distribution in the detector corresponds to the real position distribution of $\beta\beta0\nu$ events.

For comparing both methods, the reconstruction performance is investigated using the residual, which is defined as

$$r = x_{\text{prediction}} - x_{\text{true}} \ , \tag{3.1}$$

where $x_{\text{prediction}}$ is the value predicted by the reconstruction method, which is either the DNN or the EXO-200 standard analysis, and x_{true} the true Monte Carlo label of the event.

In this thesis, the standard deviation σ of the residual r of a variable x is used as a measure for the resolution

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(r_i - \mu)^2} \ , \tag{3.2}$$

where N is the number of events in the data set and μ the mean value of the data set for the residual of the variable x.

Apart from this, the correlation of the residuals of both methods, the DNN and the EXO-200 standard analysis, with respect to the MC true label is investigated. To do this, the Pearson correlation coefficient ρ is used. For two random variables X, Y it is defined as

$$\rho_{X,Y} = \frac{\mathbb{E}[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \cdot \sigma_Y} , \tag{3.3}$$

where $\mathbb{E}$ is the expectation, μ_A the mean of A and σ_A the standard deviation of A. For $\rho = +1$, the two variables X and Y are totally positive linearly correlated, $\rho = 0$ means no correlation and $\rho = -1$ is a total negative linear correlation.

3.1 Position Reconstruction

In this section, the position reconstruction of the DNN and the classical event reconstruction is investigated and compared. Firstly, the reconstruction of the single coordinates (U, V, Z) is analyzed. After that, the reconstruction of the full three-dimensional position is investigated.

The event position in low background experiments is important and is used as a measure to distinguish signal from background events. The inner detection volume is categorized as a fiducial volume because the MC simulation near the walls of the TPC does not model the electric field correctly [24]. With a better position reconstruction, more fiducial events are correctly categorized as fiducial events and vice versa. This reduces the systematic uncertainties introduced by the fiducial volume definition.

3.1.1 U-Coordinate

Firstly, the reconstruction of the U-coordinate is investigated. Figure (3.1) shows the U-coordinate reconstructed by the DNN for simulated $\beta\beta0\nu$ SS events. In the upper plot, the reconstructed U-coordinate is plotted as a function of the true label which is known in the MC. Apart from some single outliers, all reconstructed events are close to the diagonal. This shows that the DNN can reconstruct the U-coordinate with high precision.

In the lower plot, the residual of the reconstructed U-coordinate is plotted as a function of the true MC label. The residual is independent of the U-position and fluctuates around zero. This shows that no bias of the U-coordinate is introduced for different positions. Both plots use a logarithmic color scale for the number of occurrences in the bins.

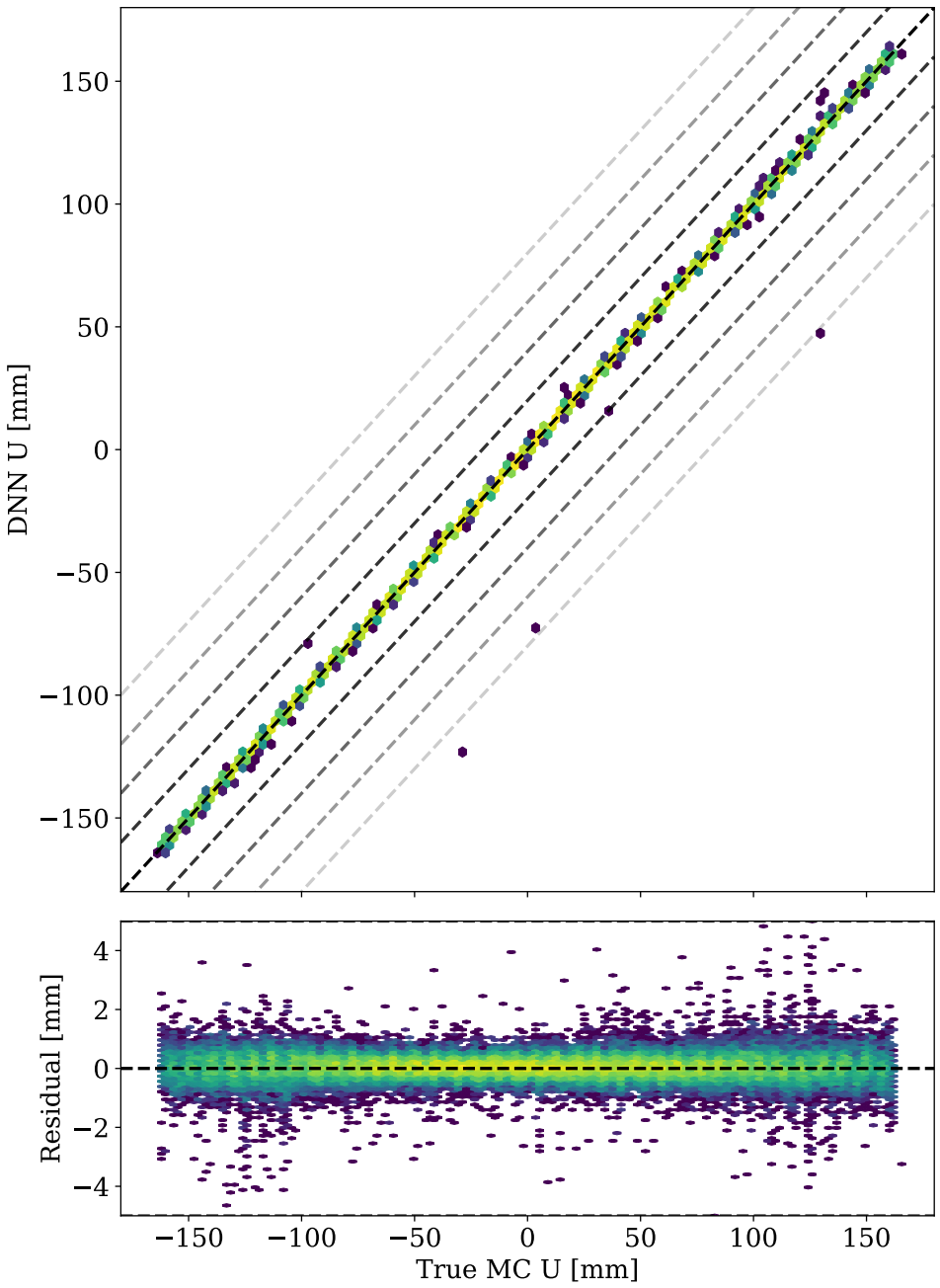

Figure 3.1: The upper plot shows the U-coordinate reconstructed by the DNN as a function of the true MC U-coordinate. The lower plot shows the residual of both values. Both plots use a logarithmic color scale to visualize the counts in each hexagonal bin.

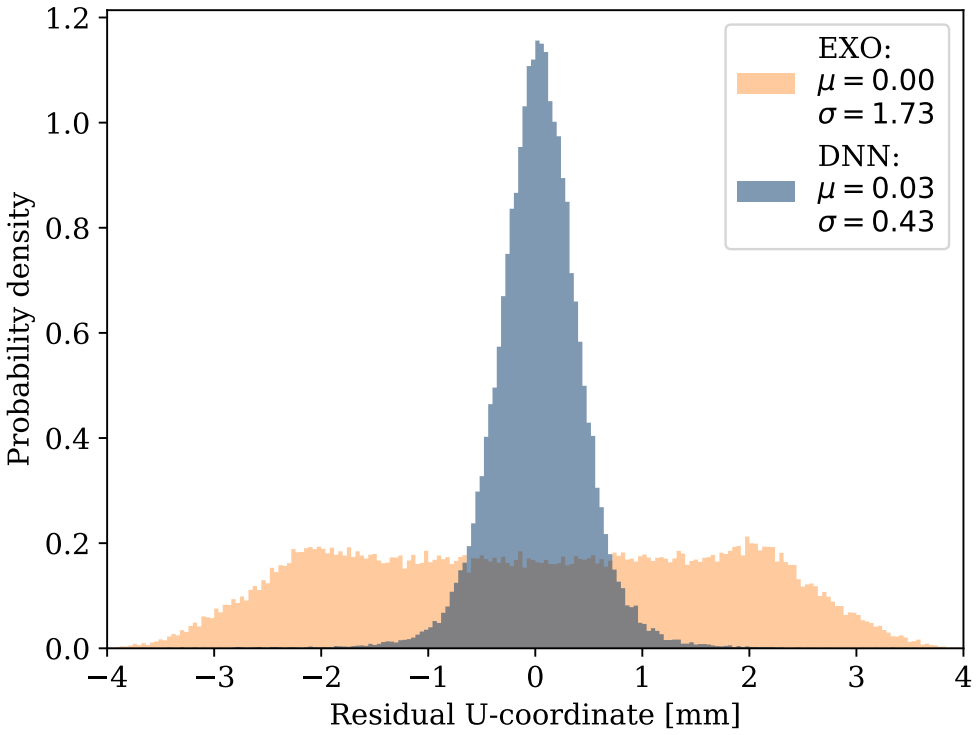

Figure 3.2: Distributions of the residual of the predicted U-coordinate for the DNN **(blue)** and the classical event reconstruction **(beige)**. The distribution is shown for $\beta\beta0\nu$ SS events.

To get an estimation of the accuracy of the DNN's reconstruction, the performance of the network is compared with the classical event reconstruction. The distributions of the residual for the reconstruction of the U-coordinate are shown in Figure (3.2). The residual of the classical event reconstruction is centered around $\mu = 0.00\,\text{mm}$ with a standard deviation of $\sigma = 1.73\,\text{mm}$. The residual of the DNN is normally distributed. The distribution is centered around $\mu = 0.03\,\text{mm}$ with a standard deviation of $\sigma = 0.43\,\text{mm}$. The residuals of both methods are symmetrical and not biased. Looking at the standard deviations, we see that the DNN clearly outperforms the standard analysis in reconstructing the U-coordinate. The resolution of the DNN is about four times better than the classical event reconstruction.

Considering the fact, that the U-wire triplets have a distance of 9 mm to each other, it is questionable whether the DNN can achieve such a good resolution.

To better understand, because of the intrinsic fluctuation the best U-resolution possible is limited through the distance between adjacent U-wire channels. Naively, the best U-coordinate resolution possible for uniformly distributed events where

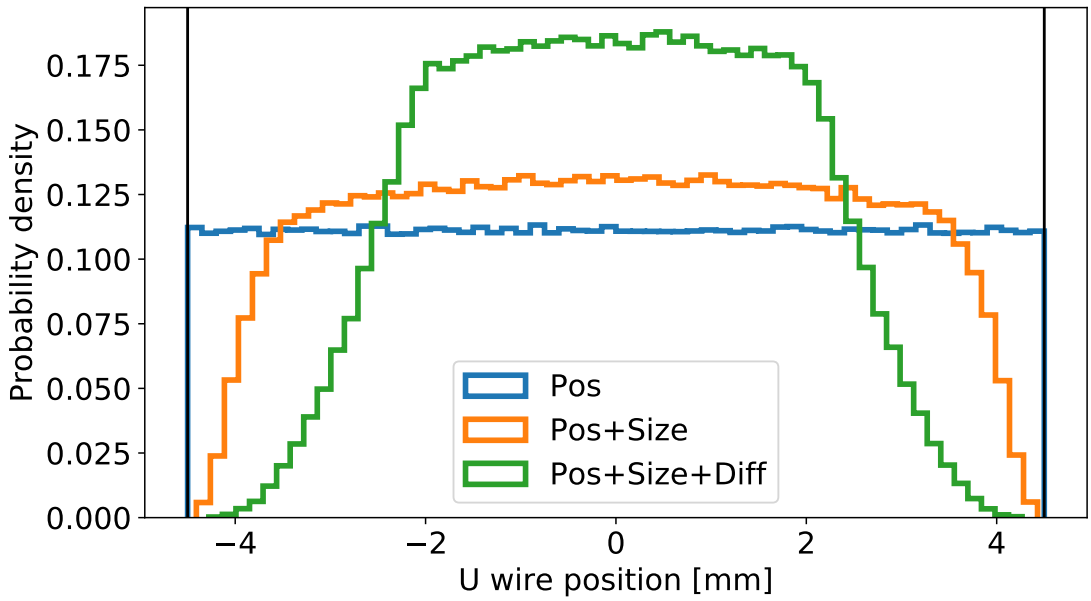

Figure 3.3: Distributions of events that are collected on one U-wire generated using a simple Monte Carlo simulation. Accounting different aspects in the MC and the drift of the secondary electrons, different event distributions are generated.

only one U-wire collects charge (one-wire event) is

$$\sigma_{\text{best}} = \frac{\text{wirepitch}}{\sqrt{12}} = \frac{9\,\text{mm}}{\sqrt{12}} \approx 2.60\,\text{mm}\,. \tag{3.4}$$

The model is valid for uniformly distributed point-shaped events, where charge is only collected on one U-wire. The U-resolution of the network for $\beta\beta0\nu$ SS one-wire events uniformly distributed in the detector is $\sigma = 0.37\,\text{mm}$. Since this is significantly better than 2.60 mm, further examinations on the U-coordinate resolution are carried out.

Using a toy Monte Carlo simulation the theoretically possible best U-coordinate resolution for different models can be approximately determined. In Figure (3.3), the distributions for the different models are plotted. Only $\beta\beta0\nu$ SS events that deposit less than 50 keV in a neighboring channel and thus are only detected in one channel are used. For this simulation, it is assumed, that only signals above this energy threshold can be reconstructed.

Since this very simple model does not fully cover the reality, the model has to be refined. Firstly, the simple model of point-like energy deposits does not correspond to reality. As an extended event would lead to a charge collection signal on the neighboring wire if the event is sufficiently large and centered between two U-wires the size of the event must be considered. Using $\beta\beta0\nu$ events and accounting for

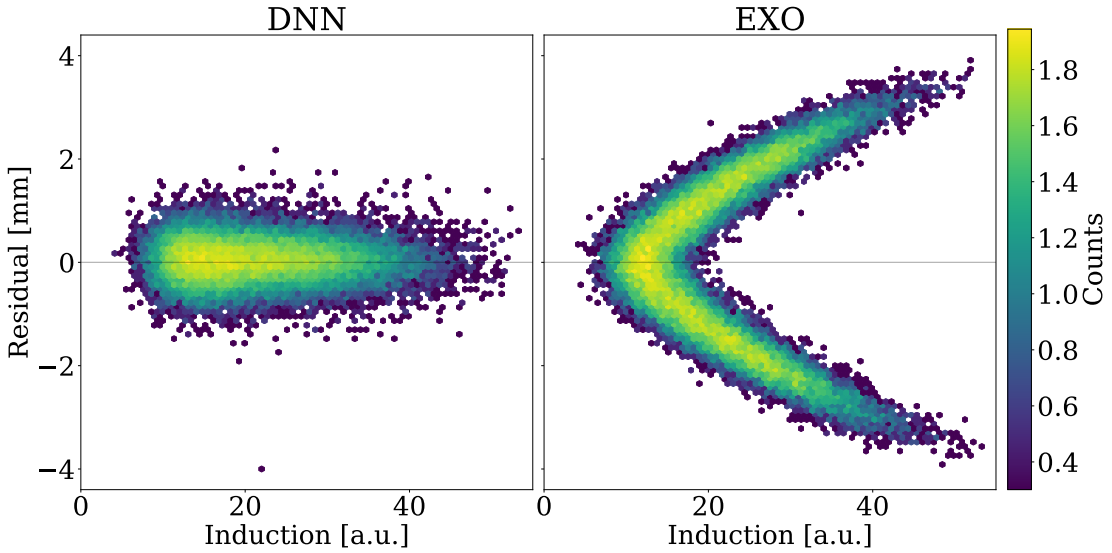

Figure 3.4: Residual of the DNN **(left)** and the classical event reconstruction **(right)** as a function of the amplitude on the neighbouring wire. As an estimator for the induction, the maximum of the neighboring channel in a temporal range of 50 µs around the maximum of the charge collection signal was used. The error on the U-coordinate reconstruction of the standard analysis gets larger with larger induction signal. The colormap is on a logarithmic scale.

the real distribution of event sizes, the best U-coordinate resolution possible is $\sigma = 2.27\,\text{mm}$. Since during the drift the size of the events is increased through diffusion, this also has to be taken into account in the model. Accounting phase I diffusion parameters for the diffusion of the events during the drift, the best possible U-coordinate resolution is found to be $\sigma = 1.67\,\text{mm}$. Since the toy Monte Carlo simulation determining this value is rather simple, this value is not the quantitative true value, but a qualitative estimator.

Since the U-coordinate resolution of the network is still significantly better, there are other features the network is using to improve the resolution. One possible feature might be an induction signal on the neighboring U-wires. The network could use the induction signal to determine an amplitude weighted mean position and thereby improve the resolution of the U-coordinate.

To check this, the residual as a function of the induction signal on the neighboring U-wires is investigated. As an estimator for the induction signal, the maximum value of both neighboring U-wires in a temporal range around the charge collection signal is used. The temporal range was $t = \pm 25\,\mu\text{s}$ around the maximum of the charge collection signal.

Figure (3.4) demonstrates the dependence between the amplitude of the neighboring channel and the residual for both reconstruction methods. There is a strong correlation between the induction signal and the residual of the EXO-200 standard analysis. For larger induction signals the absolute value of the residual of the standard analysis is rising linearly. This means that the U-coordinate reconstruction is worse with a larger induction signal. For the DNN the residual is at least constant with increasing induction signal. A larger induction signal corresponds to an event farther from the charge collecting U-wire i. e. closer to the U-wire measuring the induction signal. The EXO-200 standard analysis does not exploit the induction signal on U-wires. Because of this, it does not have the information about the relative position between the U-wires. The only information the standard analysis can use is the signal of the charge collecting U-wire. This leads to the worse U-coordinate reconstruction for increasing induction signal.

The fact, that the classical event reconstruction becomes worse but the network is unaffected with increasing induction signal, is a very good hint that the network uses this to determine the U-wire position more precisely. Using the induction signal the DNN gets a better U-coordinate resolution as the theoretically best possible value without an induction signal.

To prove, that the network uses the induction signal on the neighboring wires in order to improve the resolution in the U-coordinate beyond the limit from the toy MC, the neighboring wires of the charge collection wire are set to zero. By doing so, the network cannot use the induction as additional information.

Figure (3.5) shows the U-coordinate resolution for the standard analysis and the DNN for both, signal with induction and without induction. The beige distributions are the residual of the classical event reconstruction and the blue distributions the one of the DNN. For a better interpretation of the distributions, box plots are shown. The boxes show the lower and upper quartile of a distribution, which correspond to 25% and 75% of the cumulative distribution. The notch with the orange line in the box indicates the median. Inside the whiskers, 95% of the data are contained. The box and whisker plot does not consider the actual distribution of the data, but the ratios of the data.

Starting from the top, the first two boxes and distributions show the residual of the EXO-200 standard reconstruction and of the prediction of the network. The bottom row shows the residual of the reconstruction of the $\beta\beta0\nu$ SS one-wire events without induction signals on the neighboring U-wires. This only affects the residual of the DNN, since the classical event reconstruction does not use the induction signal on U-wires. The significantly better U-coordinate reconstruction of the DNN in the case of one-wire events with an induction signal can be seen clearly in the illustration. Without the induction signal, the DNN is approximately as good as the standard event reconstruction. This is evidence that the network exploits the

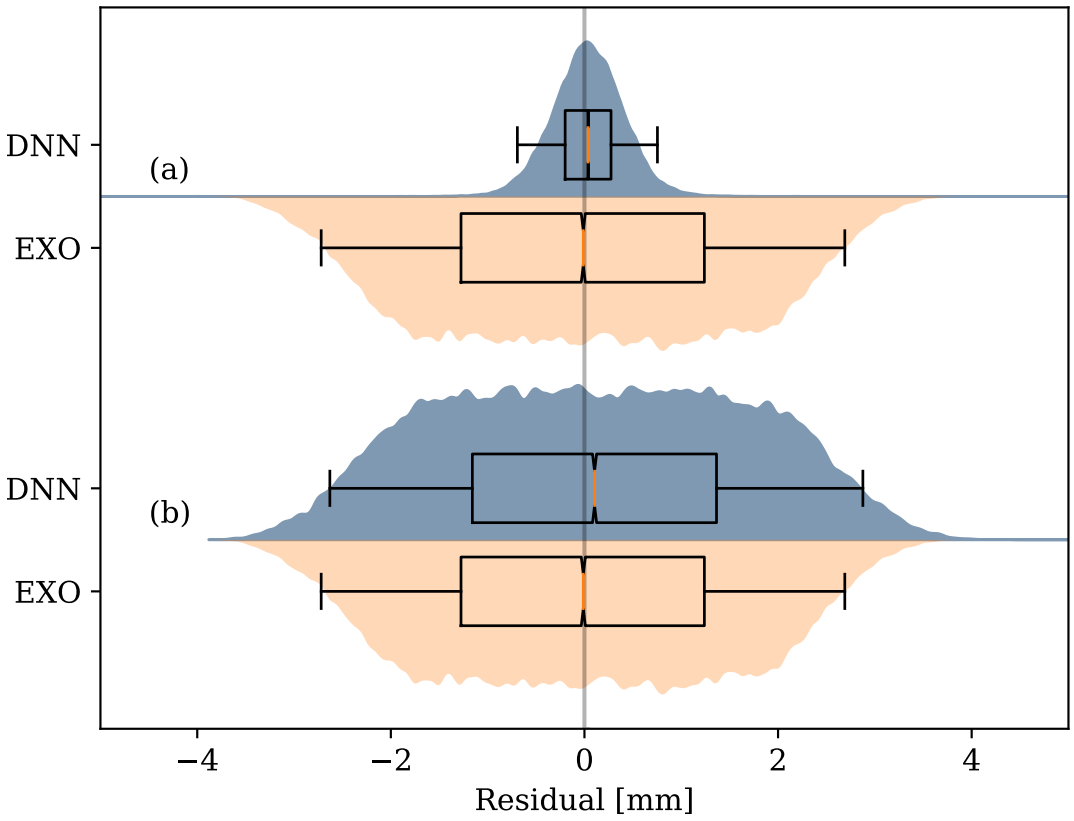

Figure 3.5: Box and violin plot of the residuals of the classical event reconstruction **(beige)** and the DNN **(blue)**. The residuals are shown for $\beta\beta0\nu$ SS one-wire events with **(a)** and without **(b)** an induction signal on the neighbouring U-wires.

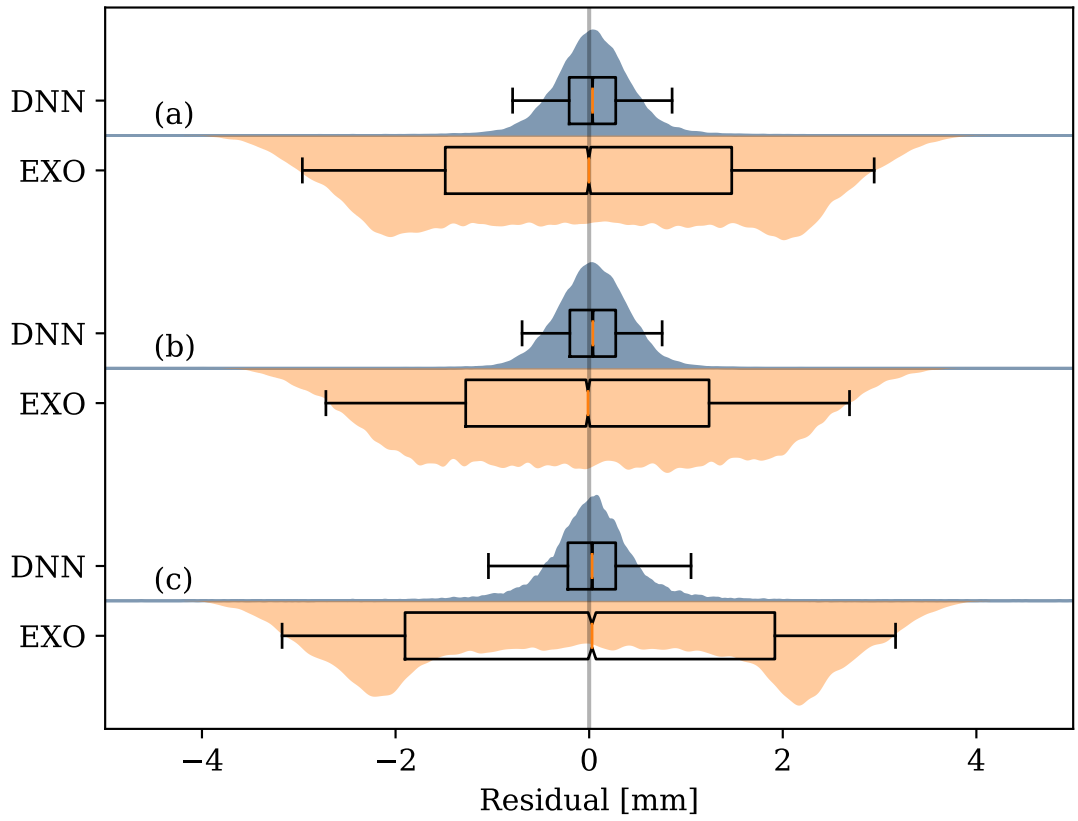

Figure 3.6: Box and violin plot of the residual of the classical event reconstruction **(beige)** and the DNN **(blue)**. The residuals are shown for all $\beta\beta0\nu$ SS events **(a)**, for $\beta\beta0\nu$ SS one-wire events **(b)**, and for $\beta\beta0\nu$ SS two-wire events **(c)**.

induction signal on the neighboring U-wires. Without the induction signal, the DNN residual has a standard deviation of 1.56 mm, which is close to the theoretical limit from the simplified Monte Carlo simulation.

As with no induction signal used by the network, the U-resolution is larger than the theoretical limit, the U-coordinate resolution of the DNN is consistent with the toy MC.

Figure (3.6) shows a box and violin plot for the residual of the U-coordinate for different subsets of $\beta\beta0\nu$ SS events. In the upper row, the distributions of the residual of all SS $\beta\beta0\nu$ events is shown. The middle row visualizes the residual for one-wire events and the third row for two-wire events.

The distributions of the residual of the DNN are approximately unchanged for all three event selections. For the two-wire events, the box is as large as for the one-wire case but with broader whiskers i. e. half of the events are reconstructed as

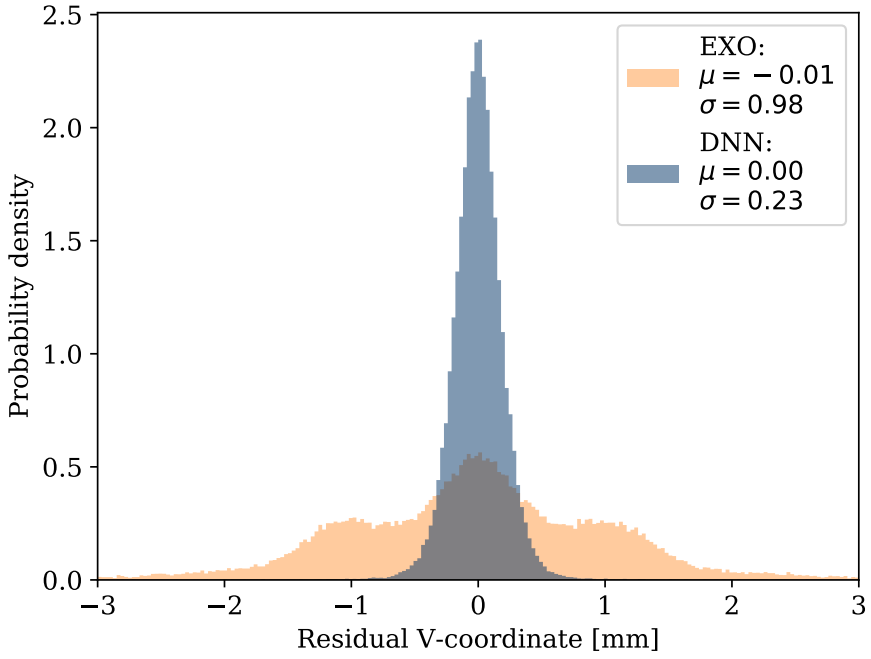

Figure 3.7: Distribution of the residual of the predicted V-coordinate for the DNN **(blue)** and the classical event reconstruction **(beige)**. The distributions are shown for $\beta\beta0\nu$ SS events.

good as the one-wire case but the other half is reconstructed worse. The EXO-200 standard analysis has a very broad distribution for SS one-wire events while for the two-wire events it has two peaks at about $\pm 2.25\,\text{mm}$. The behavior for the two-wire events is strange and should be further investigated but is not within the scope of this thesis. Comparing the DNN and the classical event reconstruction we find that the DNN is not only better for the different event selections but also has an unchanged symmetrical distribution.

3.1.2 V-Coordinate

Since the V-wires measure the charge induction signal, even for SS events, many wires measure a signal. Because of this, a better resolution than the one for the U-coordinate is expected as the position can be determined by an amplitude weighted mean position.

In Figure (3.7), the residual of the V-coordinate reconstruction is shown. The distribution of the residual of the network is unbiased, i. e. $\mu = 0.00\,\text{mm}$, and has a Gaussian shape with a standard deviation of $\sigma = 0.23\,\text{mm}$. The residual of the classical event reconstruction is unbiased, i. e. $\mu = -0.01\,\text{mm}$, and has a standard

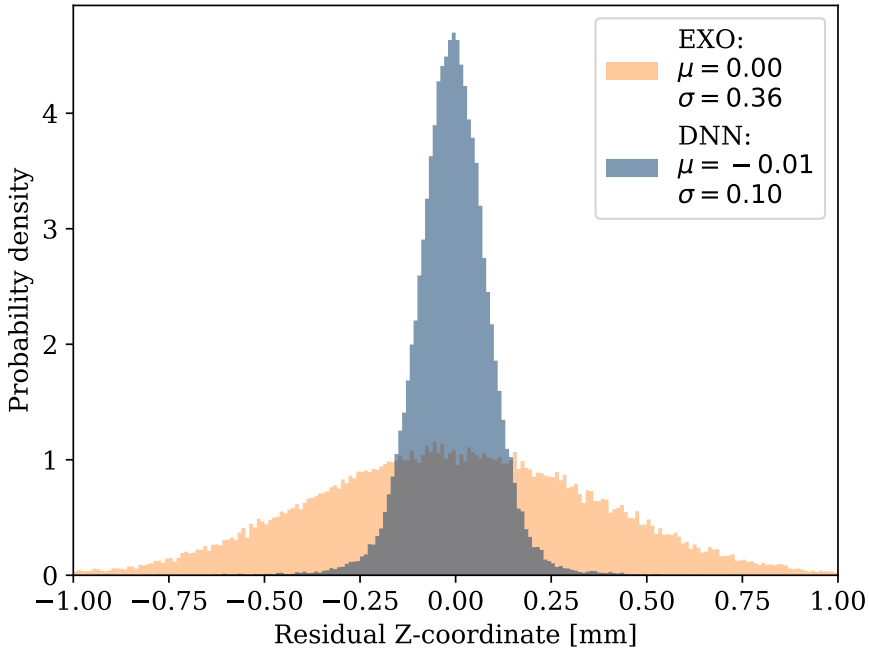

Figure 3.8: Distribution of the residual of the predicted Z-coordinate for the DNN **(blue)** and the classical event reconstruction **(beige)**. The distributions are shown for $\beta\beta0\nu$ SS events.

deviation of $\sigma = 0.98\,\text{mm}$. The DNN has a by a fourfold better V-coordinate resolution. As expected, the V-coordinate resolution is clearly better than the U-coordinate resolution.

3.1.3 Z-Coordinate

The Z-coordinate is determined by the time difference between the prompt scintillation light signals of the LAAPDs and the induction and charge collection signals on the wire planes. This means that the Z-coordinate reconstruction is only limited by the drift velocity of $0.171\,\frac{\text{cm}}{\mu\text{s}}$ and the sampling rate. Because of this, a better resolution in Z is expected than for the U- and V- coordinates.

Figure (3.8) shows the residual of both, the EXO-200 standard analysis and the DNN, for the reconstruction of the Z-coordinate. Both distributions of the residual have Gaussian shapes and are unbiased, i. e. $\mu = 0.0\,\text{mm}$. The residual of the classical event reconstruction has a standard deviation of $\sigma = 0.36\,\text{mm}$ while the DNN has a Z-coordinate resolution of $\sigma = 0.10\,\text{mm}$. As expected, the resolution in Z is better than that in U and V for both reconstruction methods. The resolution in Z of the network is almost by a fourfold better than the standard analysis.

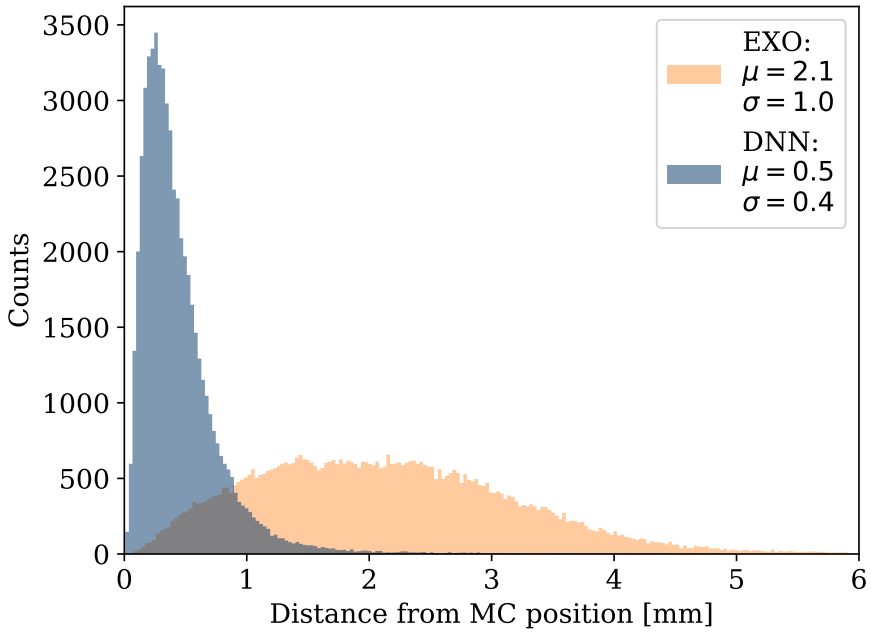

Figure 3.9: Distributions of the distance of the three-dimensional reconstructed position from the true MC position. The blue distribution is for the DNN and the beige distribution for the classical event reconstruction.

3.1.4 3D Position

Using three separately trained neural networks where each predicts either the U-, V-, or Z-coordinate, the three-dimensional position of an event can be predicted. Since the DNN outperforms the classical event reconstruction in all single coordinates, it is expected to also outperform the EXO-200 standard reconstruction in the three-dimensional position reconstruction.

To determine the euclidean distance between the true MC position and the predicted position, the reconstructed U- and V-coordinates are transformed to X- and Y-coordinates. The X-coordinate can be calculated using

$$x = \begin{cases} v - u, & z > 0 \\ u - v, & z < 0 \end{cases} \tag{3.5}$$

and the Y-coordinate

$$y = \frac{u + v}{\sqrt{3}} \,. \tag{3.6}$$

In Figure (3.9), the distance of the reconstructed three-dimensional position from the true MC position is visualized. The mean distance of the DNN predicted

position to the true MC position is $\mu = 0.5 \pm 0.4\,\text{mm}$. For the standard analysis, the mean distance to the true MC position is $\mu = 2.1 \pm 1.0\,\text{mm}$.

As expected, the DNN has a significantly better position reconstruction. The spatial resolution of the DNN is fourfold better than the EXO-200 standard analysis.

Apart from the mean of the spatial resolution, the spatial resolution should not depend on the location in the detector. Figure (3.10) shows the spatial resolution as a function of the position in the detector. The left plots show the spatial resolution for the DNN and the right plots for the classical event reconstruction. The top row shows a projection to the XZ-plane and the bottom row to the XY-plane. Only events inside the fiducial volume determined by the standard analysis are considered. Apart from statistical fluctuations and some outliers, the spatial resolution of the DNN is constant over the detector volume. For the position reconstructed by the standard analysis the spatial resolution in the middle of the detector is better than towards the walls. This can be seen in both the XY- and the XZ-plane.

As mentioned at the beginning of the section, the position reconstruction is important to distinguish fiducial events from not-fiducial events. The DNN has a fourfold better spatial resolution and its resolution is uniform over the detector. This does not fully apply to the EXO-200 reconstruction which shows worse resolution towards the walls. These two reasons are important for the experiment in order to categorize not-fiducial events correctly and to neglected them.

3.2 Energy Reconstruction

The energy resolution for low background experiments like EXO-200 is fundamentally important. A possible definition of the Region of interest (ROI) is two standard deviations from the theoretical Q-value of the $\beta\beta0\nu$ decay. A better energy resolution thus leads to a smaller region of interest and to less background in the ROI. Therefore energy reconstruction is further investigated in this section.

The energy scale reconstructed by the standard analysis E is calibrated using the Q-value of the ^{208}Tl peak of the ^{228}Th calibration source

$$E_{corr} = E \cdot \frac{Q}{\mu_{peak}}\,, \tag{3.7}$$

where E_{corr} is the calibrated energy and μ_{peak} the fitted position of the reconstructed ^{208}Tl peak. This means the reconstructed energy spectrum is scaled in a way that the reconstructed peak is at the position of the true peak.

Figure (3.11) shows the residual of the energy reconstruction for both the EXO-200 standard analysis and the DNN with respect to the MC true label. The upper plot of the figure is a histogram of the energy residual of the classical event reconstruction. The part to the right of the peak has a Gaussian-like shape.

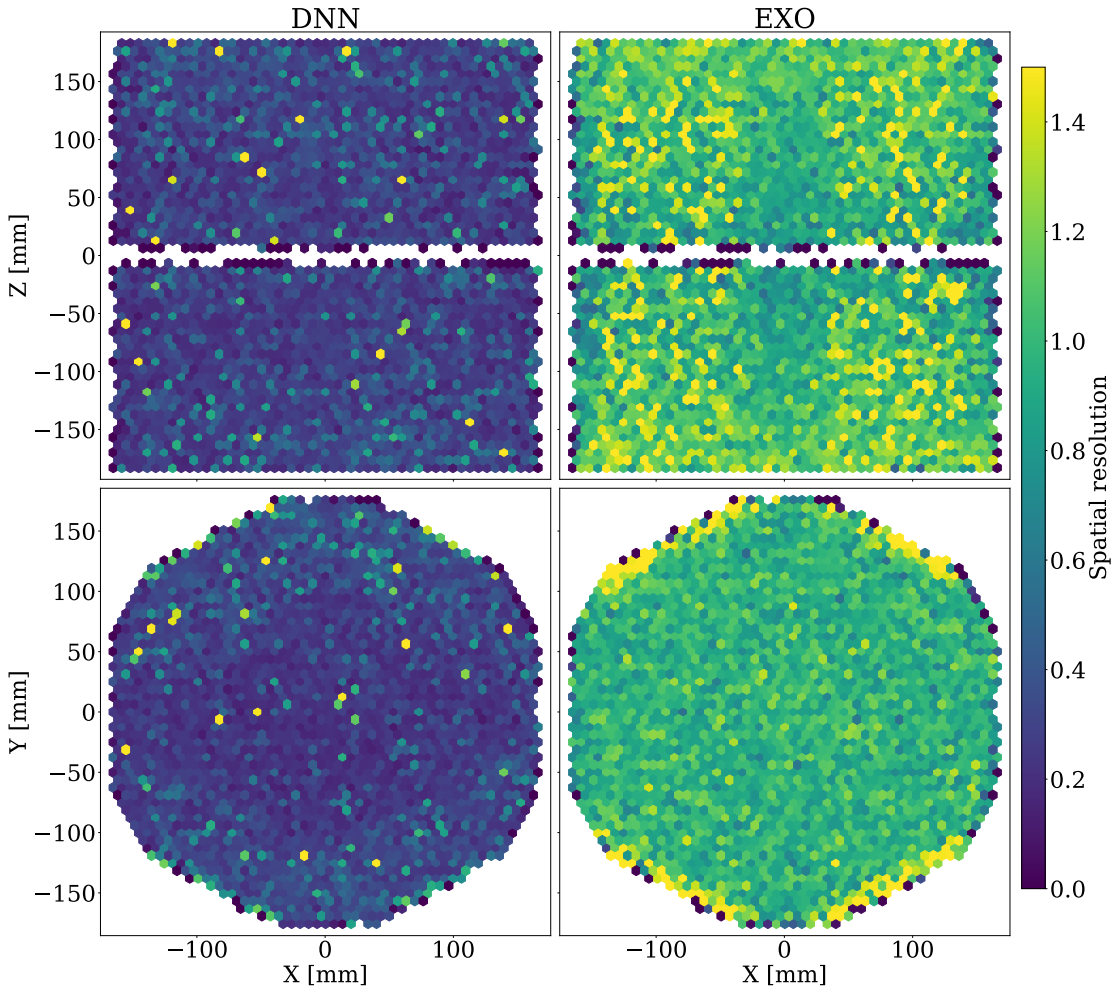

Figure 3.10: Resolution of the three-dimensional position for the DNN **(left)** and the standard event reconstruction of EXO-200 **(right)**. The top row shows the position resolution projected to the XZ-plane and the bottom row to the XY-plane.

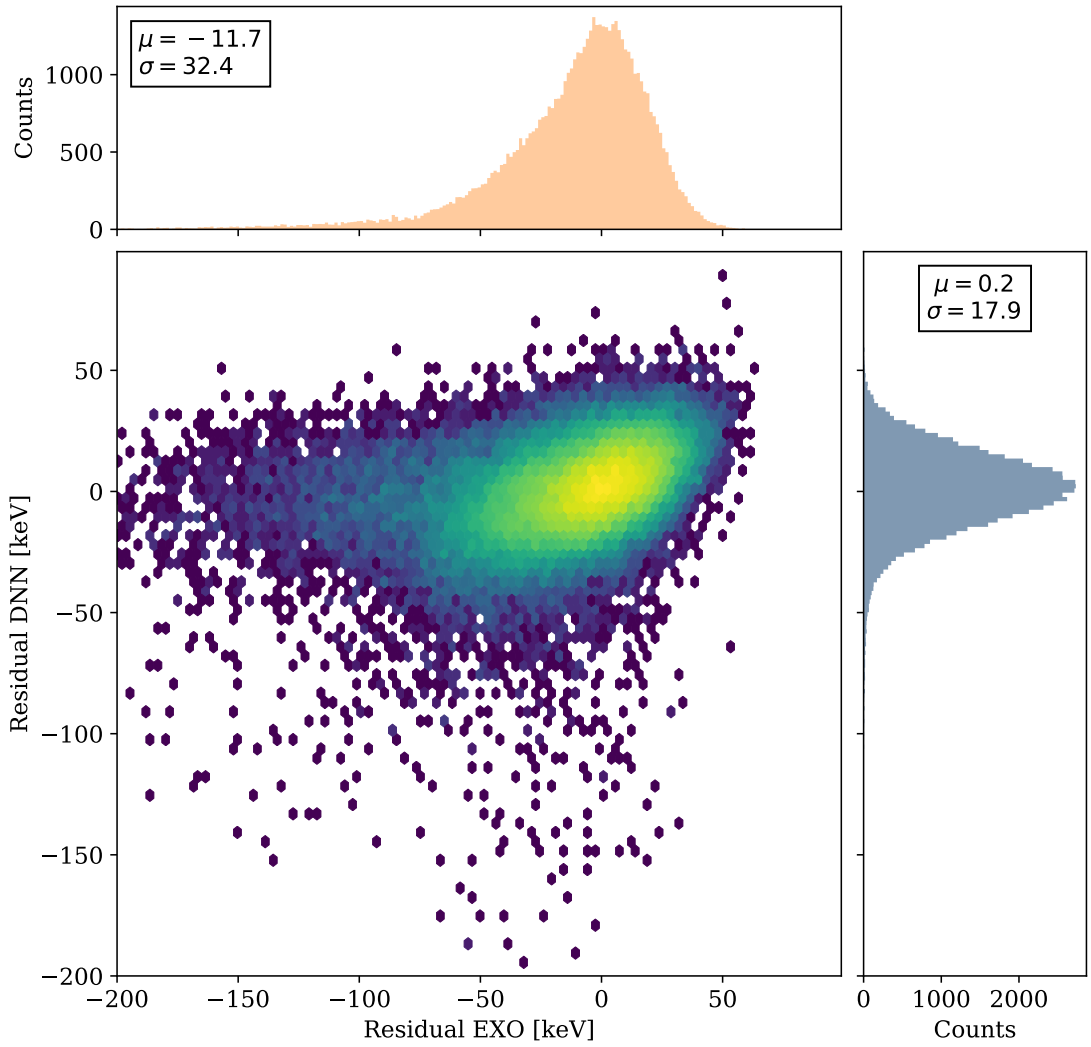

Figure 3.11: Colormap of the residual of the DNN reconstructed energy as a function of the residual of the EXO-200 standard analysis **(middle)** with respect to the MC true label. The colormap is on a logarithmic scale. The respective projection of the distribution of the residual of the DNN **(right)** and the standard analysis **(top)** is shown with its mean and standard deviation.

Whereas the part to the left of the peak has a flatter sloping flank. This means the classical event reconstruction tends to underestimate the energy of some events by failing to disentangle mixed induction and collection signals. The mean value of this method is $\mu = -11.7\,\text{keV}$ and has a standard deviation of $\sigma = 32.4\,\text{keV}$. On the right-hand side, the distribution for the DNN is shown. It has a Gaussian shape centered around $\mu = 0.2\,\text{keV}$ and a standard deviation of $\sigma = 17.9\,\text{keV}$. The resolution of the DNN is not only better but also the distribution of the residual is much more symmetric which means that it equally under- and overestimates the energy deposits.

The middle part of the figure shows the residual of the DNN as a function of the residual of the classical event reconstruction with respect to the MC true label. The colormap has a logarithmic scale. One can see that there is a positive correlation between both residuals since both methods tend to underestimate the energy of the same events and there are only very few events for which the one methods underestimates and the other one overestimates the energy or vice versa. This can also be seen looking at the Pearson correlation coefficient, which is $\rho = 0.37$. One interpretation of the correlation is that some events with the same energy deposit have a higher or lower amplitude because of the fluctuating noise level. Both reconstruction methods see these fake energy deposits and therefore tend to over- or underestimate the energy of the same events.

Not only the energy resolution is important but also it has to be uniform over the whole detector. To investigate this the energy resolution is plotted as a function of the position which is shown in Figure (3.12). The upper plots show the energy resolution in the XZ-plane and the lower plots in the XY-plane. On the right-hand side, the energy resolution of the standard analysis is visualized and on the left-hand side, the energy resolution of the DNN. Apart from fluctuations, the energy resolution is uniform over the detector volume for both reconstruction methods, which is desirable.

As mentioned at the beginning of the section the energy resolution is fundamental for low background experiments. Using the DNN to reconstruct the energy instead of the standard analysis would lead to a smaller ROI hence to less background in the ROI and better sensitivity of the experiment. Apart from this, a better energy resolution is very important to fit the probability density functions of the different components to the measurements. Additionally, the more symmetric residual of the DNN has the advantage that the background of the ^{208}Tl peak of the ^{228}Th spectrum does not leak into the ROI.

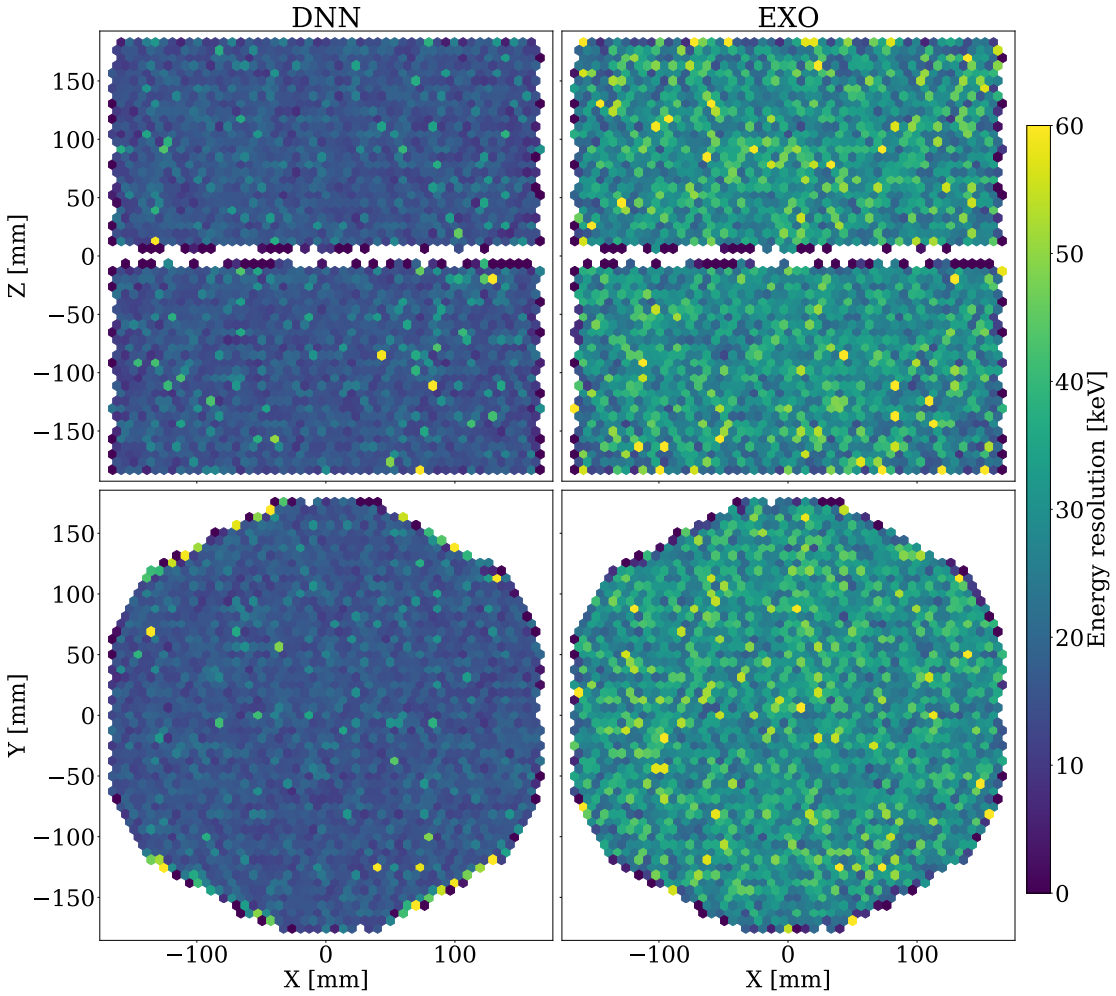

Figure 3.12: Resolution of the reconstructed energy for the DNN **(left)** and the standard event reconstruction **(right)**. The top row shows a projection of the energy resolution to the XZ-plane and the bottom row to the XY-plane.

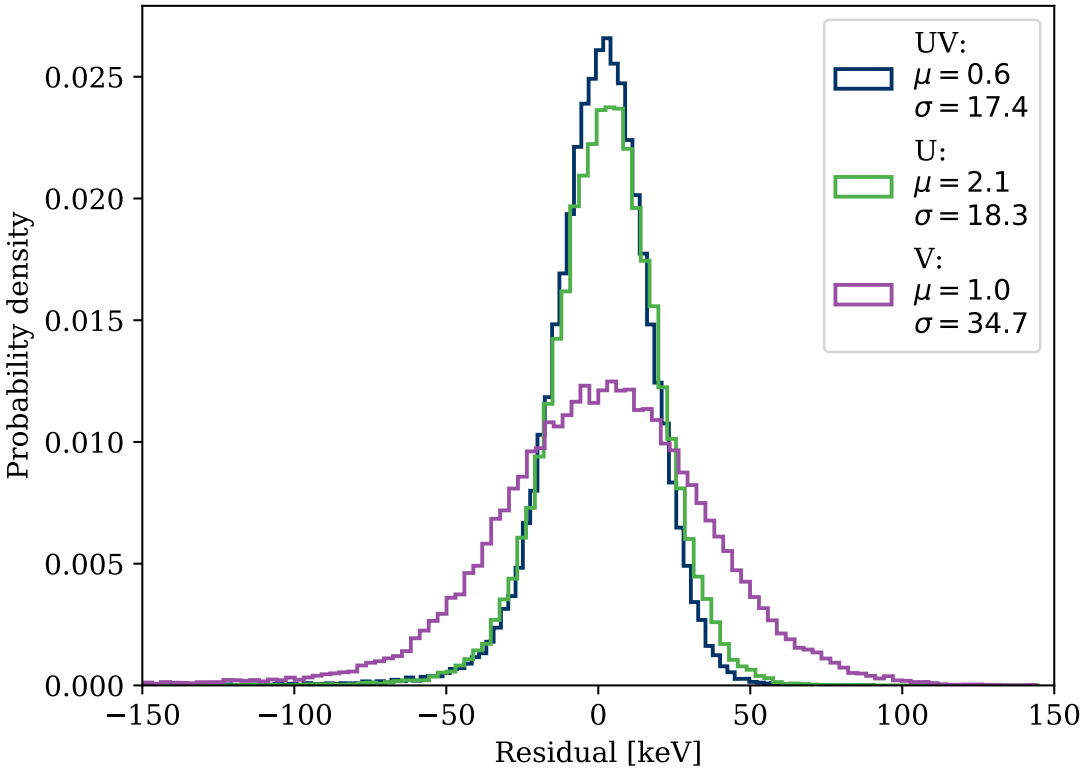

Figure 3.13: Distribution of the residual of the energy for the network trained on U- and V-wire planes **(blue)**, a network trained on only U-wires **(green)** and one trained on only V-wires **(purple)**.

3.3 Impact of U- and V-Wire Input

In this section, the impact of using both, U- and V-wire planes, compared to using only either the U- or V-wire planes as an input for the neural network is investigated. As both wire planes contain information about the energy of an event this might be redundant information that does not improve the energy resolution or might improve it because of two independent measurements with independent noise impact. For this purpose, two networks with very similar architecture as the one described in section (2.2) were trained, whereby one network only has the U-wire planes as an input and the other one only the V-wire planes. The difference to the previously used network is that the network branches with the wire planes not used are omitted.

Figure (3.13) shows the residual of the reconstructed energy for the three different networks. The residual of all three networks have Gaussian shapes centered around zero. The network using only the U-wires as an input has a significantly better

energy resolution of $\sigma = 18.2\,\text{keV}$ as the one having only the V-wires as an input, i. e. $\sigma = 34.7\,\text{keV}$. The DNN using both U- and V-wire planes as an input outperforms both networks, i. e. $\sigma = 17.4\,\text{keV}$. The difference between the network with U- and V-wire input compared to the one with only U-wire input is quite small. Nevertheless, the additional V-wire input improves the energy resolution and therefore should not be neglected especially as the energy resolution is so important.

CHAPTER 4

Reconstruction of Real Events

Contents

In this section, the performance on real data of the four networks for the U-, V-, Z-coordinate, and the energy, trained on MC simulated data, is evaluated. Since there are no true labels for the event position and the energy of real events, it is only possible to compare the performance of the DNN with the performance of the EXO-200 standard analysis. Additionally, it is possible to compare the performance of both on real data with their performance on MC data. For validation, single-site (SS) events of a ^{228}Th calibration source at position S5 are used. The secondary electrons, in reality, have a finite electron lifetime, due to impurities in the liquid xenon. In order to account for this, the reconstructed energy is corrected as explained in section (2.1.3).

4.1 Position Reconstruction

Firstly, the performance on real data of the networks reconstructing the event position is investigated. As in the previous chapter, the single coordinates are analyzed separately. In contrast to the previous chapter the three-dimensional position is not analyzed. This is due to the fact that without a true position label of the three-dimensional position a reasonable statement about the reconstruction is not possible.

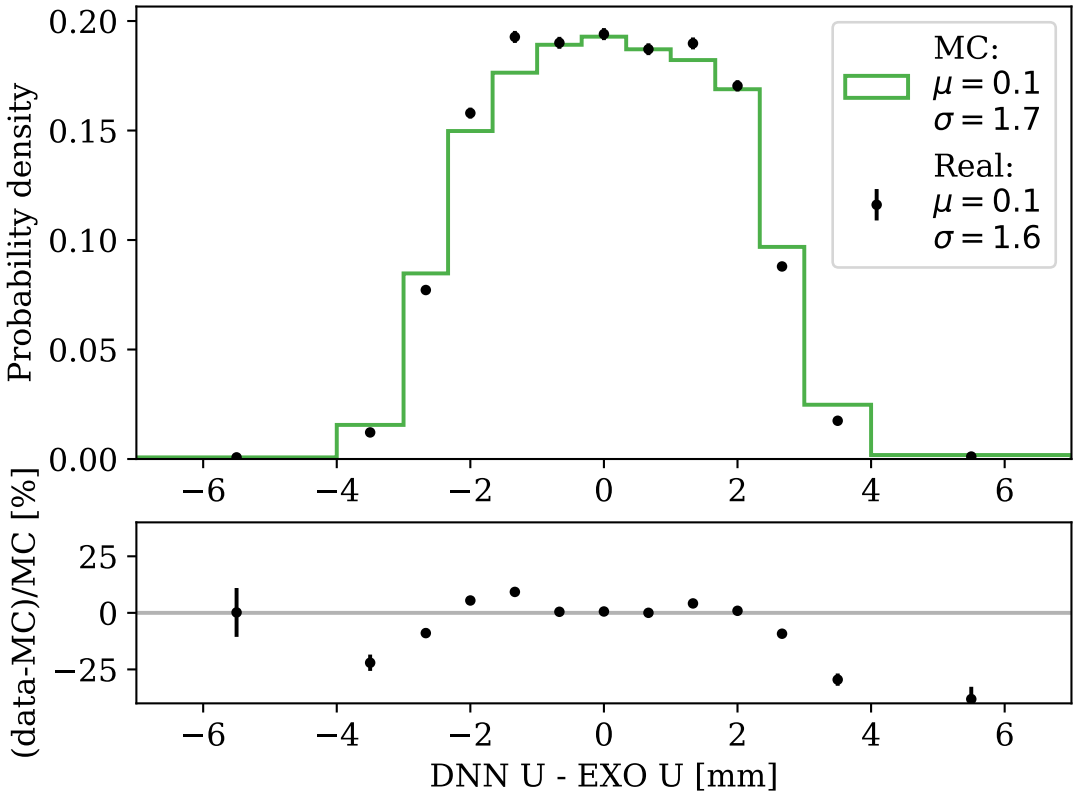

Figure 4.1: The upper plot shows the distributions of the difference between the DNN's and standard analysis' reconstructed U-coordinate. The difference is shown for real data **(black points)** and MC simulated data **(green)**. The lower plot shows the discrepancies between MC and real data.

4.1.1 U-Coordinate

Beginning with the reconstruction of the U-coordinate, the difference between the network's and standard analysis' predicted U-coordinate is investigated. The distribution of the difference is shown in Figure (4.1). Looking at the upper plot we see that for MC and real data the DNN predicts an unbiased U-coordinate $\mu = 0.1\,\text{mm}$ apart from the standard analysis' value. The standard deviation of the distribution states, that the difference between the reconstruction methods is in the mean $\sigma = 1.6\,\text{mm}$ for real data and $\sigma = 1.7\,\text{mm}$ for MC data. In the lower plot, the discrepancy between the performance on MC and on real data is shown. Apart from the bins with low statistics, the difference between the DNN and classical event reconstruction is almost the same for MC and real data. But there is a symmetric trend that with an increasing absolute value of the difference between

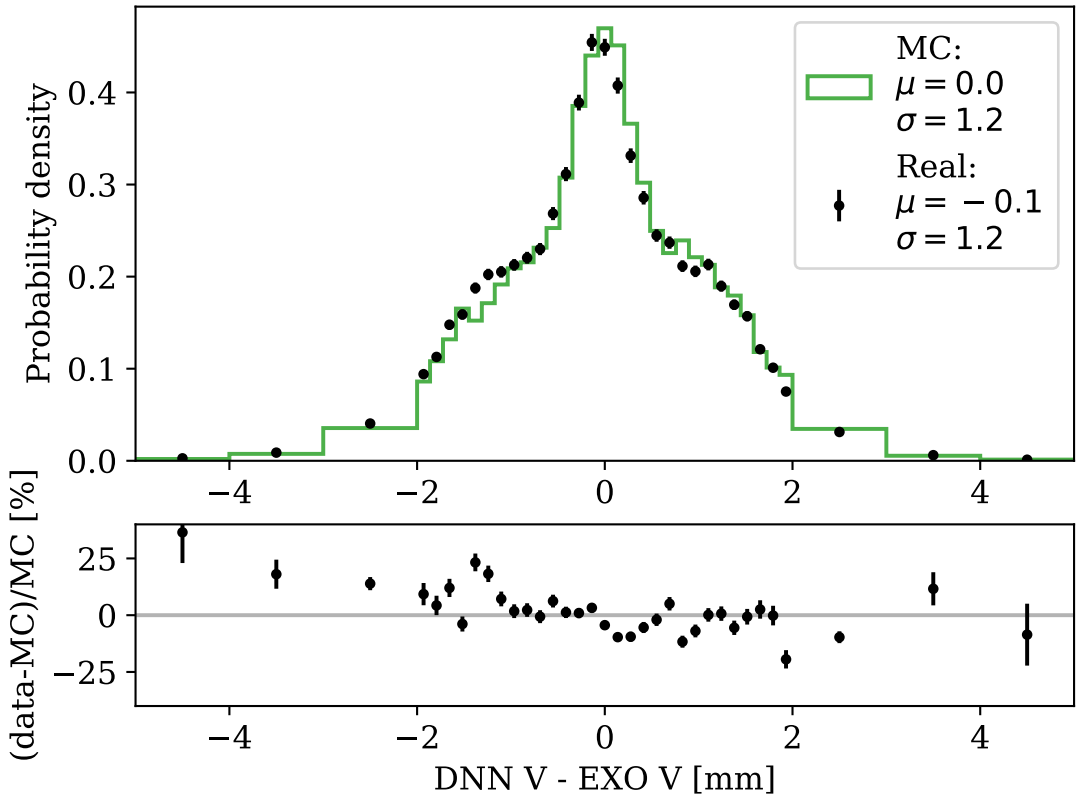

Figure 4.2: The upper plot shows the distributions of the difference between the DNN's and standard analysis' reconstructed V-coordinate. The difference is shown for real data **(black points)** and MC simulated data **(green)**. The lower plot shows the discrepancies between MC and real data.

the DNN and the classical event reconstruction the discrepancy between MC data and real data is growing.

In general, the DNN performs comparably to the standard analysis. Applying DNNs trained on MC data to real data is not always working well. A quantitative statement about the U-coordinate resolution of the DNN is not possible.

4.1.2 V-Coordinate

For the V-coordinate again the difference between the reconstruction of the DNN and the standard analysis is investigated. The difference between the two reconstruction methods is shown in Figure (4.2) for both MC and real data. Looking at the upper plot, the mean prediction for real data only differs by $\mu = -0.1\,\text{mm}$ with a deviation of $\sigma = 1.2\,\text{mm}$. Comparing this with the mean difference on MC events,

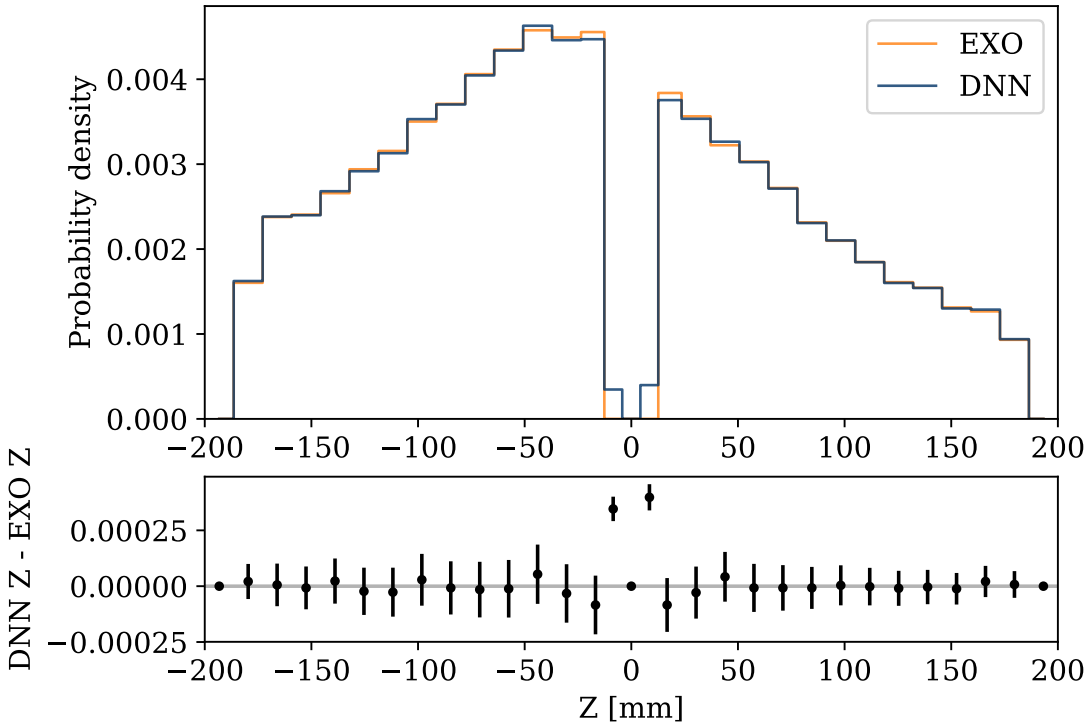

Figure 4.3: The upper plot shows the distribution of the reconstructed Z-coordinate for the DNN **(blue)** and the standard analysis **(beige)**. In the lower plot the difference between both reconstruction methods is shown.

both distributions agree. There is no mean difference between the DNN and the standard analysis for MC events, i. e. $\mu = 0.0\,\text{mm}$.

The lower plot shows the difference between the reconstruction difference between MC and real data. Apart from the very left point with low statistics, MC and real data agree within 25%. This is an indication that the differences in the performance of the DNN and the standard analysis on real data are similarly as the differences on MC data. This is not a proof but since there is no true label a quantitative statement is not possible.

4.1.3 Z-Coordinate

For the reconstruction of the Z-coordinate, the spectrum of the Z-coordinate is investigated. Figure (4.3) shows the reconstructed spectrum of the DNN and the classical event reconstruction. First of all, we see that the distributions for both reconstruction methods are very similar. The largest discrepancy is near the cathode around $Z = \pm 10\,\text{mm}$. Since the fiducial cut is made on the reconstructed

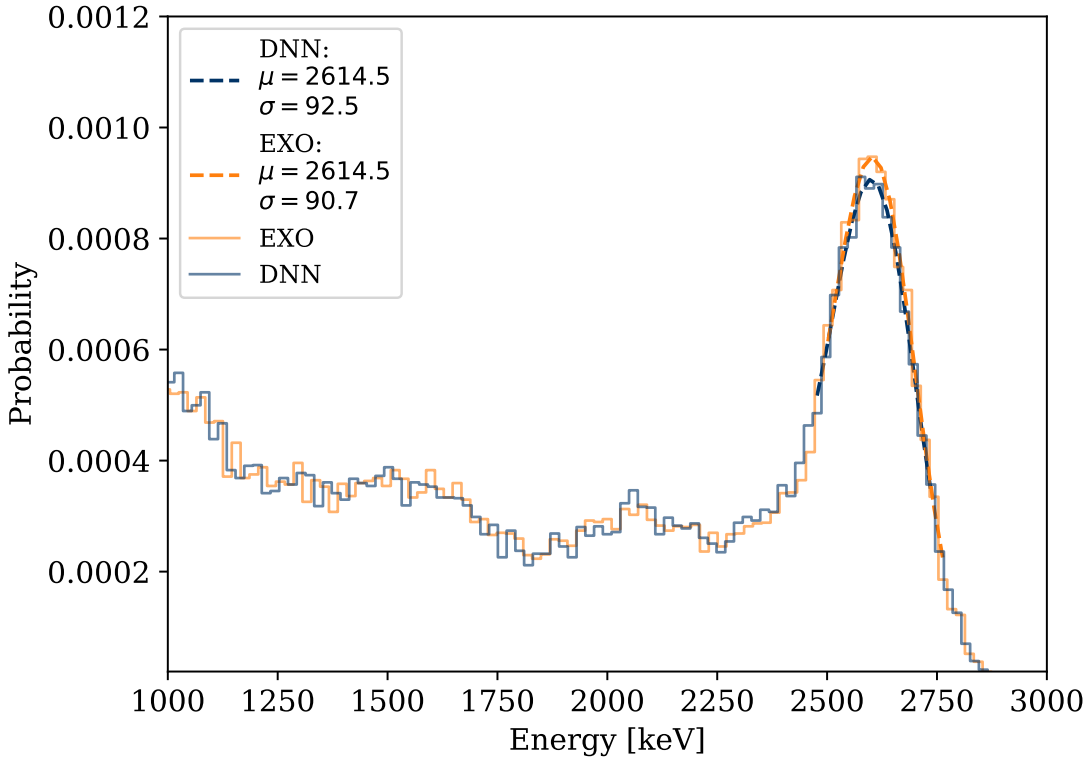

Figure 4.4: Spectrum of the reconstructed energy of the DNN **(blue)** and the standard analysis **(beige)**. For both spectra a Gaussian is fitted to the ^{208}Tl peak.

position of the standard analysis, there are no events with $|Z| < 10\,\text{mm}$ for the standard analysis. However, the DNN reconstructs some events outside the fiducial volume. These events can be seen in the first bin inside the fiducial volume for the classical event reconstruction. To have as fewer events labeled falsely as possible is why the good spatial resolution is so important. Apart from this, because of the unsymmetrical shape of the spectrum, we can see that the real position of the ^{228}Th calibration source is not equal to the nominal position $Z = 0\,\text{mm}$.

4.2 Energy Reconstruction

In this section, the network's performance on reconstructing the energy of real data is compared to the performance of the standard analysis. We compare the reconstructed spectra of the ^{228}Th calibration source at position S5. Figure (4.4) shows the spectra for both reconstruction methods. In general, both spectra look very similar. To investigate which reconstruction method performs better, we fit a Gaussian to the ^{208}Tl peak at 2614.5 keV. The DNN has a standard deviation

at the peak of $\sigma_{\mathrm{DNN}} = (3.54 \pm 0.07)\,\%$ and the classical event reconstruction of $\sigma_{\mathrm{EXO}} = (3.47 \pm 0.08)\,\%$. So the classical event reconstruction has a slightly better energy resolution. This is surprising as the DNN has a significantly better energy resolution on MC events. The big difference between the two scenarios is that the real events have a finite electron lifetime and the anti-correlation between light and charge for real data. The impact of the finite electron lifetime is corrected for real data but this correction might not have worked perfectly. Apart from this, it is a great result that the DNN works comparably to the standard analysis on real data.

Conclusion

As demonstrated in this thesis, applying Deep Learning methods can improve the analysis of physical experiments as shown for EXO-200. Comparing the performance of the network with the performance of the standard analysis of EXO-200 on simulated MC events, the DNN clearly outperforms the classical event reconstruction for all spatial coordinates and the energy. The spatial resolution can be improved by a factor of four using the DNN as it exploits the full information available in the raw signals. This includes the collection signal which is also used by the EXO-200 standard analysis but also the induction signal which improves the U-coordinate resolution significantly. The energy resolution can be improved by a factor of two compared to the classical event reconstruction. Additionally, the use of U- and V-wire planes as an input compared to only U- or V-wire planes was motivated as this improves the energy resolution.

Although there are known differences between MC simulated events and real events, the reconstruction of real events using DNNs also works well. The missing true label for real events complicates the comparison between different reconstruction methods.

A big problem for most DNN based approaches are the discrepancies between MC and real data. Having a perfect MC would solve the problem of not having a true label for real data. Minimizing this discrepancies by a better physical understanding and modelling of the detector is very hard and involves a great deal of effort.

Apart from this, new Deep Learning based ansätze are emerging. Generative Adversarial Networks (GANs) have shown to be able to refine MC events [37] or even generate MC events from scratch [38]. This is done by comparing the existing Monte Carlo data with real data and refining the MC to be indistinguishable from real data. This ansatz is currently investigated by Federico Bontempo in the EXO-200 Erlangen group.

One of the next steps is to reconstruct all energy deposits individually for MS events. Some investigations have already been done on this topic. Using a recurrent network, that is both a classifier and regressor is one possible option. The classification part predicts the number of separated energy deposits (charge clusters) in the detector, whereas the recurrent regression part predicts the energy and position of each charge cluster individually. Afterward, the information of the regression and classification part is combined and the number of clusters predicted by the classifier is kept.

Another interesting further step is using a network, that not only predicts the energy and position, but also the uncertainty of the network's prediction on the respective variable. This is possible using a specific loss function. Using this would be very useful as we would also get an estimation of how reliable a prediction is.

CHAPTER A

Bibliography

[1] Google Trends. *Deep Learning - Interesse im zeitlichen Verlauf*. URL: https://trends.google.de/trends/explore/TIMESERIES/1556103000?hl=de&tz=-120&date=2014-01-01+2019-04-17&geo=DE&q=%2Fm%2F0h1fn8h&sni=3 (visited on 17/04/2019).

[2] W. Pauli et al. *Wissenschaftlicher Briefwechsel mit Bohr, Einstein, Heisenberg u.a. Band II: 1930–1939 / Scientific Correspondence with Bohr, Einstein, Heisenberg a.o. Volume II: 1930–1939*. Sources in the history of mathematics and physical sciences. Springer Berlin Heidelberg, 1985.

[3] C. L. Cowan et al. "Detection of the Free Neutrino: a Confirmation". In: *Science* 124.3212 (July 1956), pp. 103–104. DOI: 10.1126/science.124.3212.103.

[4] KATRIN Collaboration and KATRIN Collaboration. *KATRIN design report 2004*. Tech. rep. 51.54.01; LK 01. Forschungszentrum, Karlsruhe, 2005. 245 pp. DOI: 10.5445/IR/270060419.

[5] P. S. Bhupal Dev et al. "Constraining Neutrino Mass from Neutrinoless Double Beta Decay". In: *Phys. Rev.* D88 (2013), p. 091301. DOI: 10.1103/PhysRevD.88.091301. arXiv: 1305.0056 [hep-ph].

[6] M. Goeppert-Mayer. "Double Beta-Disintegration". In: *Phys. Rev.* 48 (6 Sept. 1935), pp. 512–516. DOI: 10.1103/PhysRev.48.512.

[7] Mark G. Inghram and John H. Reynolds. "Double Beta-Decay of Te^{130}". In: *Phys. Rev.* 78 (6 June 1950), pp. 822–823. DOI: 10.1103/PhysRev.78.822.2.

[8] J. J. Gomez-Cadenas et al. "The search for neutrinoless double beta decay". In: *Rev. Mod. Phys.* 80 (Apr. 2008), pp. 481–516. DOI: 10.1103/RevModPhys.80.481..

[9] B.R. Martin and G. Shaw. *Particle Physics*. Manchester Physics Series. Wiley, 2008. ISBN: 9780470721537.

[10] W. H. Furry. "On Transition Probabilities in Double Beta-Disintegration". In: *Physical Review* 56.12 (Dec. 1939), pp. 1184–1193. DOI: 10.1103/physrev.56.1184.

[11] Werner Rodejohann. "Neutrino-less Double Beta Decay and Particle Physics". In: *Int. J. Mod. Phys.* E20 (2011), pp. 1833–1930. DOI: 10.1142/S0218301311020186. arXiv: 1106.1334 [hep-ph].

[12] J. Schechter and J. W. F. Valle. "Neutrinoless double-βdecay in SU(2)×U(1) theories". In: *Physical Review D* 25.11 (June 1982), pp. 2951–2954. DOI: 10.1103/physrevd.25.2951.

[13] Masaru Doi, Tsuneyuki Kotani and Eiichi Takasugi. "Double Beta Decay and Majorana Neutrino". In: *Progress of Theoretical Physics Supplement* 83 (1985), pp. 1–175. DOI: 10.1143/ptps.83.1.

[14] J. J. Gomez-Cadenas et al. "The search for neutrinoless double beta decay". In: *La Rivista del Nuovo Cimento* 35.2 (Jan. 2012), pp. 29–98. ISSN: 0393-697X. DOI: 10.1393/ncr/i2012-10074-9.

[15] L. Cardani. "Neutrinoless Double Beta Decay Overview". In: *SciPost Phys. Proc.* 1 (2019), p. 024. DOI: 10.21468/SciPostPhysProc.1.024. arXiv: 1810.12828 [nucl-ex].

[16] X. Qian and P. Vogel. "Neutrino mass hierarchy". In: *Progress in Particle and Nuclear Physics* 83 (July 2015), pp. 1–30. DOI: 10.1016/j.ppnp.2015.05.002.

[17] S Adrián-Martínez et al. "Letter of intent for KM3NeT 2.0". In: *Journal of Physics G: Nuclear and Particle Physics* 43.8 (June 2016), p. 084001. DOI: 10.1088/0954-3899/43/8/084001.

[18] U.D. of Energy. *WIPP - Waste Isolation Pilot Plant*. URL: http:%20//www.wipp.energy.gov/wipprecovery/recovery.html.

[19] E.-I. Esch et al. "The cosmic ray muon flux at WIPP". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 538.1-3 (Feb. 2005), pp. 516–525. DOI: 10.1016/j.nima.2004.09.005.

[20] M Auger et al. "The EXO-200 detector, part I: detector design and construction". In: *Journal of Instrumentation* 7.05 (May 2012), P05010–P05010. DOI: 10.1088/1748-0221/7/05/p05010.

[21] J. B. Albert et al. "Search for Neutrinoless Double-Beta Decay with the Upgraded EXO-200 Detector". In: *Physical Review Letters* 120.7 (Feb. 2018). DOI: 10.1103/physrevlett.120.072701.

[22] G. Plante. "The XENON100 Dark Matter Experiment: Design, Construction, Calibration and 2010 Search Results with Improved Measurement of the Scintillation Response of Liquid Xenon to Low-Energy Nuclear Recoils". 2012. URL: http://xenon.astro.columbia.edu/thesis/plante_thesis-201201.pdf.

[23] Joshua Albert. "Status and Results from the EXO Collaboration". In: *EPJ Web of Conferences* 66 (2014). Ed. by S. Lunardi et al., p. 08001. DOI: 10.1051/epjconf/20146608001.

[24] Sebastian Schmidt. "Analyses of the Electron Motion and Recombination in the EXO-200 Experiment". Oct. 2017.

[25] J. B. Albert et al. "Improved measurement of the $2\nu\beta\beta$ half-life of ^{136}Xe with the EXO-200 detector". In: *Physical Review C* 89.1 (Jan. 2014). DOI: 10.1103/physrevc.89.015502.

[26] René Vidal et al. "Mathematics of Deep Learning". In: *CoRR* abs/1712.04741 (2017). arXiv: 1712.04741. URL: http://arxiv.org/abs/1712.04741.

[27] Kurt Hornik. "Approximation capabilities of multilayer feedforward networks". In: *Neural Networks* 4.2 (1991), pp. 251–257. DOI: 10.1016/0893-6080(91)90009-t.

[28] Christopher M Bishop et al. *Neural networks for pattern recognition*. Oxford university press, 1995.

[29] Vincent Dumoulin and Francesco Visin. "A guide to convolution arithmetic for deep learning". In: *arXiv preprint arXiv:1603.07285* (2016).

[30] S. Agostinelli et al. "Geant4—a simulation toolkit". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (July 2003), pp. 250–303. DOI: 10.1016/s0168-9002(03)01368-8.

[31] W. Shockley. "Currents to Conductors Induced by a Moving Point Charge". In: *Journal of Applied Physics* 9.10 (Oct. 1938), pp. 635–636. DOI: 10.1063/1.1710367.

[32] S. Ramo. "Currents Induced by Electron Motion". In: *Proceedings of the IRE* 27.9 (Sept. 1939), pp. 584–585. DOI: 10.1109/jrproc.1939.228757.

[33] J. B. Albert et al. "Measurement of the drift velocity and transverse diffusion of electrons in liquid xenon with the EXO-200 detector". In: *Physical Review C* 95.2 (Feb. 2017). DOI: 10.1103/physrevc.95.025502.

[34] François Chollet et al. *Keras*. https://keras.io. 2015.

[35] Martın Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015. URL: https://www.tensorflow.org/.

[36] Diederik P. Kingma and Jimmy Ba. "Adam: A Method for Stochastic Optimization". In: *CoRR* abs/1412.6980 (2014). arXiv: 1412.6980.

[37] Martin Erdmann et al. "Generating and Refining Particle Detector Simulations Using the Wasserstein Distance in Adversarial Networks". In: *Computing and Software for Big Science* 2.1 (July 2018), p. 4. ISSN: 2510-2044. DOI: 10.1007/s41781-018-0008-x.

[38] Martin Erdmann, Jonas Glombitza and Thorben Quast. "Precise Simulation of Electromagnetic Calorimeter Showers Using a Wasserstein Generative Adversarial Network". In: *Computing and Software for Big Science* 3.1 (Jan. 2019), p. 4. ISSN: 2510-2044. DOI: 10.1007/s41781-018-0019-7.

[39] Luciano Maiani. "Selected topics in Majorana neutrino physics". In: *arXiv preprint arXiv:1406.5503* (2014).

CHAPTER B

List of Figures

CHAPTER C

List of Tables and Abbreviations

List of Tables

List of Abbreviations

RRZE Regionales Rechenzentrum Erlangen
ROI Region of interest
DNN deep neural network
EXO Enriched Xenon Observatory
nEXO next Enriched Xenon Observatory
TPC time projection chamber
LAAPD large area avalanche photodiodes
PCD pixelated charge deposit
ReLu rectified linear unit
SS single-site
MS multi-site
GPU graphics processing unit

Statutory Declaration

I declare that I have developed and written the enclosed Master's Thesis completely by myself, and have not used sources or means without declaration in the text. The Master's Thesis was not used in the same or in a similar version to achieve an academic grading or is being published elsewhere.

Johannes Link
Erlangen, April 26, 2019