Old Dominion University

ODU Digital Commons

Deeply Learning Deep Inelastic Scattering Kinematics

Markus Diefenthaler

Abdullah Farhat
Old Dominion University, afarh002@odu.edu

Andrii Verbytskyi

Yuesheng Xu
Old Dominion University, y1xu@odu.edu

Original Publication Citation

**THE EUROPEAN
PHYSICAL JOURNAL C**

Deeply learning deep inelastic scattering kinematics

Markus Diefenthaler[1]**, Abdullah Farhat**[2]**, Andrii Verbytskyi**[3,a]**, Yuesheng Xu**[2]

[1] Thomas Jefferson National Accelerator Facility, Newport News, VA 23606, USA
[2] Department of Mathematics and Statistics, Old Dominion University, Norfolk, VA 23529, USA
[3] Max-Planck-Institut für Physik, 80805 Munich, Germany

Abstract We study the use of deep learning techniques to reconstruct the kinematics of the neutral current deep inelastic scattering (DIS) process in electron–proton collisions. In particular, we use simulated data from the ZEUS experiment at the HERA accelerator facility, and train deep neural networks to reconstruct the kinematic variables Q^2 and x. Our approach is based on the information used in the classical construction methods, the measurements of the scattered lepton, and the hadronic final state in the detector, but is enhanced through correlations and patterns revealed with the simulated data sets. We show that, with the appropriate selection of a training set, the neural networks sufficiently surpass all classical reconstruction methods on most of the kinematic range considered. Rapid access to large samples of simulated data and the ability of neural networks to effectively extract information from large data sets, both suggest that deep learning techniques to reconstruct DIS kinematics can serve as a rigorous method to combine and outperform the classical reconstruction methods.

1 Introduction

Measurements of deep-inelastic scattering (DIS) by a multitude of experiments all-over the world [1–5] and the study of these measurements by the theoretical and experimental communities [5] have revealed information on the quark-gluon structure of nuclear matter and established quantum chromodynamics (QCD) as the theory of the strong interaction. The experiments at the HERA collider facility at the DESY research centre in Hamburg, Germany have played an important role in these studies. HERA has been the only electron–proton collider built so far [6]. The data collected in the years 1992 - 2007 have provided truly unique informa-

tion on the internal structure of the proton and other hadrons [7].

The key component in these studies has been a precise reconstruction of the DIS kinematics, using information from the accelerator and the detectors. Multiple methods have been applied at the HERA experiments [8] to reach an optimal precision in each particular measurement. Each of the classical reconstruction methods uses only partial information from the DIS event and is subject to specific limitations, either arising from the detector or the assumptions used in the method.

With the DIS measurements at the upcoming Electron-Ion Collider in mind [9], we present in this work a novel method for the reconstruction of DIS kinematics based on supervised machine learning and study its performance using Monte Carlo simulated data from the ZEUS experiment [10] at HERA. For our approach, we develop deep neural network (DNN) models that are optimised for the problem and are allowed to take full information from the DIS event into account. We train the DNN models on simulated data from the ZEUS experiment and compare the results from our trained model with the results from the classical reconstruction methods.

We show that the reconstruction of the DIS kinematics using deep neural networks provides a rigorous, data-driven method to combine and outperform classical reconstruction methods over a wide kinematic range. In the past, neural networks had already been used in the context of DIS experiments [11] and we expect that our novel method and similar approaches will play an even more important role in ongoing and future DIS experiments [12,13].

2 Deep inelastic scattering

Deep inelastic scattering is a process in which a high-energy lepton (l) scatters off a nucleon or nucleus target (h) with

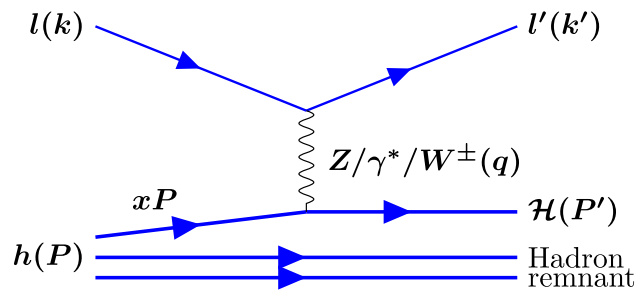a e-mail: andrii.verbytskyi@mpp.mpg.de (corresponding author)

Fig. 1 Schematic representation of Deep Inelastic Scattering process at Born level

large momentum transfer (the momentum of each entity is given in parenthesis):

$$l(k) + h(P) \rightarrow l'(k') + \mathcal{H}(P') + \text{remnant}. \tag{1}$$

The detectors in collider experiments are designed to measure the final state of the DIS process, consisting of the scattered lepton l' and the hadronic final state (HFS) $\mathcal{H}$. The latter consists of hadrons with a relatively long lifetime as well as some photons and leptons but does not include the hadron remnant. The H1 [14] and ZEUS experiments were not able to register the remnant of the target due to its proximity to the proton beam pipe.

2.1 Deep inelastic scattering at Born level

In the leading order (Born) approximation, the leptons interact with quarks in the hadrons by the exchange of a single virtual γ or Z boson in the neutral current (NC) reaction, and the exchange of single $W^{\pm}$ boson in the charged current (CC) reaction. The kinematics of the leading order DIS process in a Feynman diagram-like form is shown in Fig. 1. In this paper, we will only consider the neutral current electron scattering off a proton in a collider experiment. In this reaction, the final state lepton is a charged particle (electron or positron) that can be easily registered and identified in the detector.

With a fixed centre-of-mass energy, $\sqrt{s} = \sqrt{(k+P)^2}$, two independent, Lorentz-invariant, scalar quantities are sufficient to describe the deep inelastic scattering event kinematics in the Born approximation. Typically, the used quantities are:

– the negative squared four-momentum of the exchanged electroweak gauge boson:

$$Q^2 = -q \cdot q = -(k - k')^2, \tag{2}$$

– the Bjorken scaling variable, interpreted in the frame of a fast moving nucleon as the fraction of incoming nucleon

longitudinal momentum carried by the struck parton:

$$x = \frac{Q^2}{2P \cdot q}. \tag{3}$$

In addition to that, the inelasticity y is used to define the kinematic region of interest. It is defined as the fraction of incoming electron energy taken by the exchanged boson in the proton rest frame

$$y = \frac{P \cdot q}{P \cdot k}. \tag{4}$$

Therefore, for the DIS an equation

$$Q^2 = syx \tag{5}$$

holds. However, the Born-level picture of the DIS process is not sufficient for the description of the observed physics phenomena. A realistic description of DIS requires the inclusion of higher order QED and QCD processes [15].

2.2 Higher order corrections to deep-inelastic scattering process

The DIS process with leading order QED corrections can be written as

$$l(k) + h(P) \rightarrow l'(k') + \gamma(P_{\gamma}) + X(P') + \text{remnant}, \tag{6}$$

so the kinematics is defined not only by the kinematic of the scattered electron and the struck parton, but also by the momentum of the radiated photon, P_{γ}. The lowest order electroweak radiative corrections can be depicted in a form of Feynman diagrams as shown in Fig. 2a–d and should be considered together with the virtual corrections e–g. The Fig. 2a, d correspond to the initial state radiation (ISR) and the Fig. 2b, c to the final state radiation (FSR).

With the virtual corrections taken into account in the DIS process, Eq. (2) no longer holds, i.e.

$$q \cdot q \neq -(k - k')^2. \tag{7}$$

The presence of higher order QCD processes, (e.g. the boson-gluon fusion in Fig. 2h and QCD Compton Fig. 2i, j makes the kinematic description of the DIS process even more complicated. Therefore, the exact definitions of the kinematic observables used in the analysis of the DIS events and the corresponding simulations are essential for the correct physics interpretation.

2.3 The simulation of the DIS events and the kinematic variables in the simulated events

The simulation of the inclusive DIS events in the Monte Carlo event generators (MCEGs) starts from the simulation at the parton level, i.e. the simulation of the hard scattering process
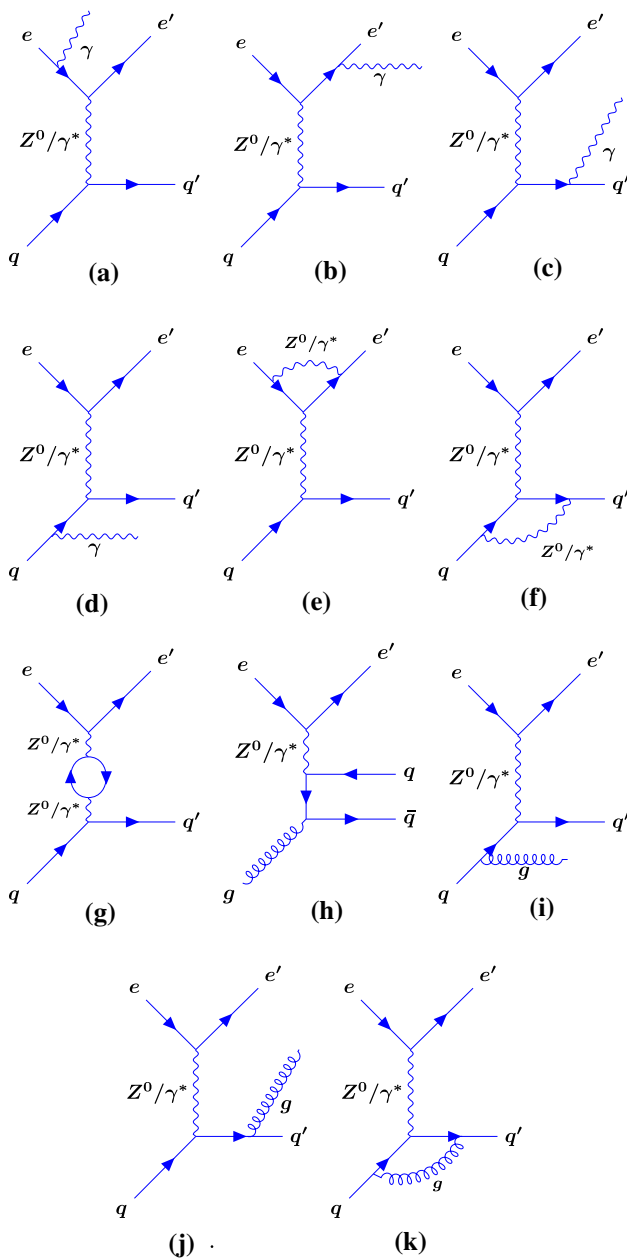
Fig. 2 Feynman diagrams for Deep Inelastic Scattering process with some leading order electroweak corrections (**a**–**g**) and QCD corrections (**h**–**k**). The proton remnant is omitted

and the kinematics of the involved partons, e.g. given by parton distribution functions (PDFs) [16] for the given hadron and considering processes with all types of partons in the initial state. The modelling of the hard scattering process combines the calculations of the perturbative QED and/or QCD matrix elements for the $2 \rightarrow n$ processes at parton level with the different QCD parton cascade algorithms designed to take into account at least some parts of the higher order perturbative QCD corrections not present in the calculations of matrix elements.

The simulated collision events on the particle (hadron) level are obtained using the parton level simulations as input and applying phenomenological hadronisation and particle decay models to them.

As of 2022, multiple MCEG programs are capable of simulating the inclusive DIS process at the hadron level with different levels of theory precision and sophistication of modelling of hadronisation, beam remnant and parton cascades, e.g. `Pythia6` [17], `Pythia8` [18], `SHERPA-MC` [19], `WHIZARD` [20] and `Herwig7` [21]. In addition to that, the `Lepto` [22], `Ariadne` [23], `Cascade` [24] and `Rapgap` [25] programs can simulate the DIS process using parts of the `Pythia6` framework for the simulation of hadronisation processes and decays of particles.

As it was discussed above, DIS beyond the Born approximation has a complicated structure which involves QCD and QED corrections [9]. The most recent MCEG programs, e.g. `Herwig7`, `Pythia8`, `SHERPA-MC`, or `WHIZARD`, contain these corrections as a particular case of their own general-purpose frameworks or are able to use specialised packages like `OpenLoops` [26], `blackhat` [27] or `MadGraph` [28]. In general, the modern MCEGs do not specify their definitions of the DIS kinematic observables, but in some cases they can be calculated from the kinematics of the initial and final state, both for the true and reconstructed kinematics. For instance, under some assumptions, the Q^2 in the event could be calculated according to Eq. (2). The total elimination of the ambiguities for such calculations is not possible, as the final state kinematics even at the parton level depend on the kinematics of all the emitted partons. The calculations of the kinematic observables from the momenta of particles at hadron level add an additional ambiguity related to the identification of the scattered lepton and the distinction of that lepton from the leptons produced in the hadronisation and decay processes.

Contrary to the approaches adopted in modern MCEG programs, the MCEGs used for the HERA experiments relied upon generator-by-generator implementations of the higher order QED and QCD corrections specific for DIS or alternatively applied `HERACLES` [15] for the corrections.

The way to get a simulation of the DIS collision even in a specific detector is the same as for any other type of particle collision event. It involves simulation of the particle transport through the detector material, simulation of detector response and is typically performed in `Geant` [29,30] or similar tools. The simulated detector response is passed to the experiment-specific reconstruction programs and should be indistinguishable from the real data recorded by the detector and processed in the same way.

2.4 Reconstruction of the kinematic variables at the detector level

The kinematics of the DIS events are reconstructed in collider experiments by identifying and measuring the momentum of the scattered lepton l' and/or the measurements of the hadronic final state ($\mathcal{H}$). The identification of the scattered lepton is ambiguous even at the particle level of the simulated DIS collision events. The same ambiguity is present in the reconstructed real and simulated data at the detector level. Therefore, the identification of the scattered electron candidate for the purposes of physics analyses is a complicated task on itself and was a subject of multiple studies in the past, some of which also involved neural network-related techniques [11]. In our paper, we rely on the standard method of the electron identification at the ZEUS experiment [11] and discuss solely the reconstruction of the kinematic variables using the identified electron and other quantities measured in the detector.

The physics analyses performed in the experiments at HERA relied on the following quantities for the calculation of the kinematic observables x and Q^2:

- The energy ($E_{l'}$) and polar angle ($\theta_{l'}$) of the scattered electron. Most of the DIS experiments are equipped to register the scattered electron using the tracking and calorimeter detector subsystems. While the tracking system is able to provide information on the momentum of the scattered electron, the calorimeter system can be used to estimate the energy of the electron and the total energy of the collinear radiation emitted by the electron. At the detector level, the estimation of these energies can be done by comparing the momentum of the electron reconstructed in the tracking system and the energy deposits registered in the calorimeter system around the extrapolated path of the electron in the calorimeter.
- The energy of the HFS expressed in terms of the following convenient variables:

$$\delta_{\mathcal{H}} = \sum_{i \in \mathcal{H}} E_i - P_{Z,i} \tag{8}$$

and

$$P_{T,\mathcal{H}} = \sqrt{\left(\sum_{i \in \mathcal{H}} P_{X,i}\right) + \left(\sum_{i \in \mathcal{H}} P_{Y,i}\right)}, \tag{9}$$

where the sums run over the registered objects i excluding the scattered electron. Depending on the analysis requirements, the used objects could be either registered tracks, energy deposits in the calorimeter system, or a combination of both.

The measurements of the quantities listed above overconstrain the reconstruction of the DIS kinematics. Therefore, in the simplest case, any subset of two observables in $E_{l'}$, $\theta_{l'}$, $\delta_{\mathcal{H}}$, and $P_{T,\mathcal{H}}$, can be used for the reconstruction.

In our analysis, we consider three specific classical reconstruction methods based on these observables which were used by the ZEUS collaboration in the past: the electron, the Jacques-Blondel, and the double-angle methods. We briefly provide some details on the methods in this section, while a more detailed description can be found elsewhere [31].

The electron (EL) method uses only measurements of the scattered lepton, $E_{l'}$ and $\theta_{l'}$, to do the reconstruction of Q^2 and x. The kinematic variables calculated from these measurements are given by:

$$Q^2_{EL} = 2E_l E_{l'}(1 + \cos\theta_{l'}) \tag{10}$$

and

$$x_{EL} = \frac{E_l E_{l'}(1 + \cos\theta_{l'})}{E_P(2E_l - E_{l'}(1 - \cos\theta_{l'}))}. \tag{11}$$

The electron method provides precise reconstruction of kinematics, but, since it uses only information from the scattered lepton, this method is affected by initial and final state QED radiation. Namely, the QED radiation registered in the detector separately from the scattered electron will not be taken into account in the calculations with this method. Practically, the reconstruction with this method gives reasonable results when E_l and $E_{l'}$ are significantly different from one another, but the resolution and stability becomes poor otherwise.

The Jacques-Blondel (JB) method uses only measurements of the final state hadronic system, $\delta_{\mathcal{H}}$ and $P_{T,\mathcal{H}}$, for the reconstruction. The kinematic variables are calculated from these by:

$$Q^2_{JB} = \frac{2E_l P^2_{T,\mathcal{H}}}{2E_l - \delta_{\mathcal{H}}}, \tag{12}$$

$$x_{JB} = \frac{2E_l Q^2_{JB}}{s\delta_{\mathcal{H}}}. \tag{13}$$

The JB method is resistant to possible biases because of unaccounted QED FSR, but requires precise measurements of the particles momenta in the hadronic final state. The factors that limit the precision of the measurements are the uncertainties in the particle identification, the finite resolution of the calorimeter and tracking detectors, the inefficiencies of these detectors, the impossibility of the particle detection around the beampipe, and the presence of objects that avoid detection (e.g. neutrinos from particle decays).

The double-angle (DA) method combines measurements from the scattered lepton and the final-state hadronic system, $\theta_{l'}$ and $\gamma_{\mathcal{H}}$, to perform the kinematic reconstruction as

follows:

$$Q_{DA}^2 = \frac{4E_l^2 \sin \gamma_{\mathcal{H}}(1 + \cos \theta_{l'})}{\sin \gamma_{\mathcal{H}} + \sin \theta_{l'} - \sin(\gamma_{\mathcal{H}} + \theta_{l'})}, \tag{14}$$

$$x_{DA} = \frac{E_l \sin \gamma_{\mathcal{H}}(1 + \cos \theta_{l'})}{E_P \sin \theta_{l'}(1 - \cos \gamma_{\mathcal{H}})}, \tag{15}$$

where the angle $\gamma_{\mathcal{H}}$ is defined as

$$\cos \gamma_{\mathcal{H}} = \frac{P_{T,\mathcal{H}}^2 - \delta_{\mathcal{H}}^2}{P_{T,\mathcal{H}}^2 + \delta_{\mathcal{H}}^2}. \tag{16}$$

The angle $\gamma_{\mathcal{H}}$ depends on the ratio of the measured quantities $\delta_{\mathcal{H}}$ and $P_{T,\mathcal{H}}$, and thus, uncertainties in the hadronic energy measurement tend to cancel, leading to good stability of the reconstructed kinematic variables. Similar to the electron method, when E_l and $E_{l'}$ are significantly different from one another, the double-angle method provides reliable results, but the resolution and stability are poor otherwise.

2.5 The methodology of measurements in the deep inelastic ep collisions

The methodology of measurements at lepton-hadron colliders in general is similar to the methodologies used at e^+e^- and hadron-hadron colliders. Briefly, in the most cases the quantity of interest is measured from the real data registered in the detector corrected for detector effects. The corrections are estimated by comparing the analysis at the detector level with the same analysis at particle level using detailed simulations of the collision events with the inclusion of higher order QED and QCD processes.

The main difference between the measurements at lepton-hadron colliders and elsewhere, is in the way the measurements involve collision kinematics. At e^+e^- colliders, the initial kinematics of the interactions is given by the lepton energies that are known parameters of the accelerators. Therefore, it is straightforward for most of the measurements in the e^+e^- experiments to estimate the centre-of-mass energy of the hard collision process. In hadron-hadron collider experiments, there is no way to measure the kinematic properties of the partons initiating the collision process, as the involved partons cannot be observed in a free state and most measurements in the hadron-hadron collisions are inclusive in the kinematics of the initial state. The DIS collisions at electron–proton colliders take a middle stance between these cases. The kinematic observables of the DIS process are measured on an event-by-event basis at the detector level using the methods described above.

In an experiment, the measurements of event kinematics is affected by various effects. For a proper comparison of the measurements of HFS, e.g. jet cross-sections or event shape observables to corresponding perturbative QCD (pQCD) predictions, the detector-level measurements are unfolded for

detector effects while hadronisation correction factors are calculated using MCEGs or specialised programs and applied to the pQCD predictions [32–35]. The prescription for calculation of those correction factors vary depending on the HFS quantities measured and the used definitions of the kinematic observables. Typically, at ZEUS and other experiments at HERA, after the unfolding of the detector effects, the measurements were also scaled by radiation correction factors to facilitate a comparison to theoretical calculations available at Born level in QED, see e.g. Ref. [36]. The factors were obtained from separate high-statistics MC simulations. This is a well-understood Monte Carlo approach and our deep learning technique can be used with it in exactly the same way as the classical methods, both for the experimental and the MC data. For future DIS measurements, e.g. at the upcoming Electron-Ion Collider, it is expected that the effects of QED and QCD radiation can be treated in an unified formalism [35], with the QED effects taken into account into a factorised approach. Our DNN-based reconstruction of DIS kinematics is compatible with such a factorised approach as well.

Therefore, in this analysis, we keep the calculation of correction factors for QED and hadronisation effects out of scope and limit the discussion to detector-level measurements. At particle-level in the MC generated events, we use a single definition of "true" kinematic observables that is based on the kinematics of the exchanged boson extracted from the MC event information.

Namely, we use the definitions of the true-level observables Q_{true}^2 and x_{true} that follow the definitions implemented in ZEUS Common NTuples [37,38] and were used in many ZEUS analyses. With these definitions the Q_{true}^2 is calculated directly from the squared four-momentum of the exchange boson q_{boson},

$$Q_{\text{true}}^2 = -|q_{\text{boson}}|^2. \tag{17}$$

The x_{true} is calculated according to the formula

$$x_{\text{true}} = \frac{Q_{\text{true}}^2}{y_{\text{after ISR}} s_{\text{after ISR}}}, \tag{18}$$

with

$$y_{\text{after ISR}} = \frac{q_{\text{boson}} \cdot P_{\text{proton beam}}}{(P_{\text{lepton beam}} - P_{\text{ISR radiation}}) \cdot P_{\text{proton beam}}} \tag{19}$$

and

$$s_{\text{after ISR}} = |P_{\text{proton beam}} + P_{\text{lepton beam}} - P_{\text{ISR radiation}}|^2, \tag{20}$$

where $P_{\text{proton beam}}$, $P_{\text{lepton beam}}$ and $P_{\text{ISR radiation}}$ represent the four-momenta of the proton beam particles, lepton beam particles and the momenta lost by the lepton beam particles to ISR. Thus, $y_{\text{after ISR}}$ corresponds to the fraction of energy of

the bare lepton (i.e. without the ISR) transferred to the HFS in the centre of mass of the proton and $s_{\text{after ISR}}$ to the centre-of-mass-energy squared of the proton and the incoming bare electron.

3 Machine learning methods

The reconstruction of DIS event kinematics is overconstrained by the previously mentioned measurements, $E_{l'}$, $\theta_{l'}$, $\delta_{\mathcal{H}}$, and $P_{T,\mathcal{H}}$. We trained an ensemble of deep neural networks (DNNs) to reconstruct x and Q^2 by correcting results from classical reconstruction methods using the information on the scattered lepton and the final-state hadronic system.

The ensemble DNN method presented here is a new approach designed specifically to address this reconstruction problem. The remainder of this section discusses in detail specifics of the DNN architectures, the specific optimisation methods used to find the optimal parameters defining the DNNs, and the specific structure of the ensemble models. Shown rigorously in what follows, the main motivation for using the unique model is:

– the universal approximation capabilities of the DNN models with our specific architectures
– the necessary reduction in the approximation error with the increase in the depth of our DNN models
– avoidance of a degradation problem [39] in training due to the residual structure of our DNN models.

Further details on the DNN-based reconstruction method will be provided in Ref. [40].

3.1 Neural networks architecture

The problem of reconstruction of DIS event kinematics can be posed as a task to construct a functional relationship between a set of input variables x and a set of response variables y. In many cases, such a relation is a continuous function, and can be approximated to arbitrary accuracy with a neural network. The corresponding neural network can have various architectures. The simplest architecture is a sequence of fully-connected, feed-forward (hidden) layers. For an input $x \in \mathbb{R}^m$, the output of the first hidden layer is:

$$h^1(x) = \alpha(A_0 x), \quad h^1 \in \mathbb{R}^{d_1}, \tag{21}$$

where $A_0 : \mathbb{R}^m \to \mathbb{R}^{d_1}$ is an affine transformation, the composition of a translation and a linear mapping, and $\alpha : \mathbb{R} \to \mathbb{R}$ is a nonlinear function, called an activation function. Define the application of α on a vector $x := (x_1, x_2, ..., x_m)^T \in \mathbb{R}^m$, by $\alpha(x) := (\alpha(x_1), \alpha(x_2), ..., \alpha(x_m))^T$.

For $A_1 : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$, the output h^1 is passed into a second hidden layer:

$$h^2(x) = \alpha\left(A_1 h^1(x)\right), \quad h^2 \in \mathbb{R}^{d_2}, \tag{22}$$

and so on, for a desired number of hidden layers. An affine transformation on the output of the final hidden layer is the output of the network. For $i \in \mathbb{N}$, call d_i the number of nodes in hidden layer i.

Define $\mathbf{\Psi}^D_{m,n}(\alpha)$ to be the set of neural networks with D hidden layers that map $\mathbb{R}^m \to \mathbb{R}^n$ with continuous, nonlinear activation function α. If $\psi \in \mathbf{\Psi}^D_{m,n}$, then there exists a set of natural numbers $\{d_i\}_{i=0}^{D+1} \subset \mathbb{N}$ with $d_0 = m$ and $d_{D+1} = n$, a set affine transformations $\{A_i : \mathbb{R}^{d_i} \to \mathbb{R}^{d_{i+1}}\}_{i=0}^{D}$, such that, for $x \in \mathbb{R}^m$,

$$\begin{cases} h^0(x) = x, \\ h^{i+1}(x) = \alpha\left(A_i h^i(x)\right) \text{ for } 0 \leq i < D, \\ \psi(x) = A_D h^D(x). \end{cases} \tag{23}$$

Specific universality properties of $\mathbf{\Psi}^D_{m,n}(\alpha)$ are proven in Refs. [41,42]. In particular, there exists some element in the class of neural networks that is arbitrarily close to any continuous function. For this reason, we say that $\mathbf{\Psi}^D_{m,n}(\alpha)$ is dense in the space of continuous functions.

We consider a particular subclass of neural networks that also maintain the universal approximation property. We call it $\mathbf{\Phi}^D_{m,n}(\alpha)$. If a function $\phi \in \mathbf{\Phi}^D_{m,n}(\alpha)$, there exists a set of natural numbers $\{d_i\}_{i=0}^{D+1} \subset \mathbb{N}$ with $d_0 = m$ and $d_{D+1} = n$, a set of affine transformations $\{A_i : \mathbb{R}^{d_i} \to \mathbb{R}^{d_{i+1}}\}_{i=0}^{D-1}$, a set of matrices $\{W_i \in \mathbb{R}^{d_i \times n}\}_{i=0}^{D}$, such that, for $x \in \mathbb{R}^m$,

$$\begin{cases} h^1(x) = \alpha(A_0 x), \\ h^{i+1}(x) = \alpha\left(A_i h^i(x)\right) \text{ for } 0 \leq i < D, \\ \phi(x) = W_0 x + \sum_{i=1}^{D} W_i h^i(x). \end{cases} \tag{24}$$

Here the affine transformations $A_i, i = 0, 1, \ldots, D-1$, have the form

$$A_i x = M_i x + b_i, \quad \text{for } x \in \mathbb{R}^{d_i}$$

for matrices $M_i \in R^{d_i \times d_{i+1}}$ and vectors $b_i \in R^{d_{i+1}}$. The function ϕ defined by (24) is determined by matrices M_i and W_i, and vectors b_i whose entries/components will be embraced in a single notation ω later. That is, $\phi = \phi_\omega$. When ω is chosen via an optimisation problem, $\phi(x)$ is a function of the input variable $x \in \mathbb{R}^{d_0}$.

Let $\mathbf{\Phi}^D_{m,n}(\alpha)$ be the function class in which we search for an optimal kinematic reconstruction method. In addition to the universality property, it can be shown that this class is good for searching for an appropriate reconstruction function because there the necessary reduction in the approximation error with the increase in the depth and the residual structure avoids a degradation problem in the weights during training [40].

3.2 Optimisation methods

In any given problem to be solved with the DNN, the task is to choose an optimal function from the class $\mathcal{F} = \mathbf{\Phi}^D_{m,n}(\alpha)$ to be a functional relationship between a set of input variables x and a set of response variables y.

Each function $\phi \in \mathcal{F}$ is completely determined by the elements of the matrices and vectors that determine the values of the hidden layers. Let $\omega \in \mathbb{R}^p$ be the collection of all these elements, where p is the total number of them. We can then write for any $x \in \mathbb{R}^m$: $\phi(x) = \phi_\omega(x)$.

Choosing an optimal function from this class entails finding the collection of parameters ω in $\mathbb{R}^p$ that minimises $\mathcal{L}_\mathcal{D}$, the expected discrepancy or generalisation error between y and $\phi_\omega(x)$, over the joint probability distribution of x and y, $\mathcal{D}$, defined by:

$$\mathcal{L}_\mathcal{D}(\omega) = \mathbb{E}_{x,y \sim \mathcal{D}} \ell(\omega, x, y) = \int \ell(\omega, x, y) d\mathcal{D}(x, y), \quad (25)$$

where ℓ is some loss function measuring the discrepancy between y and $\phi_\omega(x)$.

With a randomly sampled data set $\{x_i, y_i\}_{i=1}^N$, if N is sufficiently large, then the generalisation error can be approximated by an empirical error:

$$\mathcal{L}_\mathcal{D} \approx \frac{1}{N} \sum_{i=1}^N \ell(\omega, x_i, y_i). \quad (26)$$

This summation is sometimes called the fidelity term, as it measures the discrepancy between the data and model predictions. A commonly used fidelity term is the mean square error, with:

$$\ell(\omega, x_i, y_i) = \|y_i - \phi_\omega(x_i)\|_2^2. \quad (27)$$

The fidelity term used in this study is a modification of this, the mean square logarithmic error:

$$\ell(\omega, x_i, y_i) = \|\log(y_i) - \log(\phi_\omega(x_i))\|_2^2. \quad (28)$$

Due to the universality of neural networks, there is a model that can achieve zero empirical error. However, in the presence of noise this means the model can overfit to the data sample and loose its generalisability. This problem can be addressed by adding a regularisation term to the optimisation problem as a penalty for certain irregular behaviour. Common regularisation terms include the ℓ^2-norm, used to limit the size of the parameters, and the ℓ^1-norm, which can induce sparse solutions. For out analysis, we choose the ℓ^1-regularisation. The Theorem 1.3 in [43] and Proposition 27 of [44] provide an evidence that minimising the ℓ^1 norm provides sparse optimal solutions with a minimal number of nonzero elements. Well-constructed models should be able to generalise the information from one given sample to any possible event. Thus, regularisation with the ℓ^1 norm produces a

model determined by a minimal number of parameters so that the optimal solution does not fit completely to the training set and loose its generalisability.

Therefore, the determination of the optimal neural network model consists of minimising this final loss function, the sum of the sample fidelity term and a weighted regularisation term:

$$\min_{\omega \in \mathbb{R}^p} \frac{1}{N} \sum_{i=1}^N \ell(\omega, x_i, y_i) + \mathcal{R} \cdot ||\omega||_1, \quad (29)$$

where ℓ is defined in Eq. (28) and $\mathcal{R}$ is a hyperparameter to determine the magnitude of the regularisation.

Expression 29 can be minimised using stochastic gradient methods on batches of the data sample [45]. The training is accelerated using classical momentum methods [46]. In particular, randomly select an initial set of parameters ω^0. Select a sequence of step sizes, or learning rates, $\mathcal{L}^k$ that diminish to zero. Randomly selecting a batch of data with indices $I \subset \{1, ..., N\}$. Choosing a momentum parameter μ. By defining:

$$v^{k+1} = \mu v^k - \mathcal{L}^k \nabla_\omega L_I(\omega^k), \quad (30)$$

$$\omega^{k+1} = \omega^k + v^{k+1}. \quad (31)$$

Then the sequence ω^k converges to a set of parameters defining the optimal neural network with the minimal generalisation error, in the sense described here.

3.3 The model

We construct a model to rigorously weight classically derived reconstructions of x and Q^2 with corrections based on measurements from the final state lepton and hadronic system. The final reconstruction of these observables with the neural network approach are labelled below as Q^2_{NN} and x_{NN} respectively.

The constructed model is an ensemble of networks from the previously defined function class $\mathbf{\Phi}^D_{m,1}(\alpha)$, where α is the rectified linear unit (ReLU). The values of D and m vary with each network in the ensemble.

The NC DIS events studied in this analysis are by definition the events containing the scattered electron in the final state, therefore we aim to reconstruct the Q^2_{NN} primarily from the related observables, i.e. using the properties of the electron directly measured in the experiment. In particular, we use as inputs three set of variables: the classically reconstructed kinematic observables $\left(Q^2_{EL}, Q^2_{DA}, Q^2_{JB}\right)$, measurements on the scattered lepton $(E_{l'}, \theta_{l'})$, and measurements on the final-state hadronic system $(\delta_\mathcal{H}, P_{T,\mathcal{H}})$. We reconstruct the Q^2 in the form:

$$Q^2_{NN} = A_{Q^2}\left(Q^2_{EL}, Q^2_{DA}, Q^2_{JB}\right) + L_{Q^2}\left(A_{Q^2}, E_{l'}, \theta_{l'}\right)$$
$$+ H_{Q^2}\left(A_{Q^2}, \delta_\mathcal{H}, P_{T,\mathcal{H}}\right), \quad (32)$$

in which A_{Q^2} could be understood as a rigorous average of classically derived reconstructions, L_{Q^2} is a correction term computed from the scattered lepton, and H_{Q^2} is another correction term computed from the final-state hadronic system. In our analysis A_{Q^2}, L_{Q^2}, H_{Q^2} are simultaneously trained networks in $\Phi_{3,1}^5(\alpha)$, with each hidden layer of the networks containing 2000 nodes.

The x observable for the NC DIS events is actually calculated from the electron-related observables as well. Therefore, we reconstruct x_{NN}, with Q_{NN}^2 also as an input, for a total of eight inputs ($x_{EL}, x_{DA}, x_{JB}, E_{l'}, \theta_{l'}, \delta_{\mathcal{H}}, P_{T,\mathcal{H}}, Q_{NN}^2$), in the form:

$$x_{NN} = A_x \left(x_{EL}, x_{DA}, x_{JB} \right) + L_x \left(A_x, Q_{NN}^2, E_{l'}, \theta_{l'} \right) \\ + H_x \left(A_x, Q_{NN}^2, \delta_{\mathcal{H}}, P_{T,\mathcal{H}} \right), \tag{33}$$

where A_x, L_x, and H_x are defined similarly to A_{Q^2}, L_{Q^2}, and H_{Q^2}, but A_x is a network in $\Phi_{3,1}^{20}(\alpha)$ with each hidden layer containing 1000 nodes, and L_x, H_x are networks in $\Phi_{4,1}^{10}(\alpha)$, with each hidden layer containing 500 nodes.

The number of hidden layers and the number of nodes in each hidden layer were each progressively increased until sufficiently desirable results were achieved. Smaller networks provided good results on average, but larger networks were needed to find best results in small, specific regions of the kinematic space. A further increasing of the numbers of hidden layers and nodes per layer beyond the chosen values was found to not significantly alter the performance of the kinematic reconstruction, due to the convergence results of deep neural networks. The convergence theorems of deep neural networks with the ReLU activation function as the number of layers increases were recently established in [47,48].

In the ensemble neural network model, we used the ReLU

$$\alpha(x) := \max(x, 0), \quad x \in \mathbb{R} \tag{34}$$

as the nonlinear activation function It has been shown [49] that with a gradient descent algorithm, using the ReLU function as the activation function provides a smaller training time compared to that with the use of functions with saturating nonlinearities, such as a sigmoid or hyperbolic tangent function. The reduced training time enabled us to experiment with more sophisticated networks. In addition to this, the ReLU functions do not need input normalisation to prevent them from saturating, which is a desirable property for the present analysis. Moreover, it was shown in [47] that the selection of the ReLU activation function produces a neural network as a piecewise linear function over a nonuniform partition, of the domain, determined by parameters involved in the affine transformations A_i. Such a structure ensures a good representability of the neural network and can overcome the problem of underfitting.

To accommodate for the large range of the Q^2 and x variables in the analysis, we select the loss function defined in Eq. (28) for the training of the DNN models.

4 Experimental setup

The Monte Carlo simulated events used to train our deep neural networks were specifically generated using the conditions of $e^{\pm}p$ scattering in the ZEUS detector at HERA. A detailed description of the ZEUS detector can be found elsewhere [10].

In our analysis, we have used two samples of Monte Carlo simulated e^+p DIS events that are provided by the ZEUS collaboration. These samples were generated with an inclusion of QED and higher order QCD radiative effects using the HERACLES 4.6.6 [15] package with DJANGOH 1.6 [50] interface and the ARIADNE 4.12 and LEPTO 6.5.1 packages for the simulation of the parton cascade. For both samples the same set of kinematic cuts was applied during the generation, the same set of PDFs were used, CTEQ5D [51] and the same hadronisation settings were used to model the hadronisation with the Pythia6 program. Therefore, the essential difference between the two samples is the way the higher order corrections are partially modelled with the corresponding algorithms (QCD cascades). Namely, the LEPTO MCEG utilises the parton shower approach [52], while ARIADNE implements a color-dipole model [53]. Accordingly, we label the data-sets produced by the LEPTO generator as "CDM" data sets and those with ARIADNE as "MEPS" data sets.

The generated particle-level events were passed through the ZEUS detector and trigger simulation programs based on Geant 3.21 [29], assuming the running conditions of the ZEUS experiment in the year 2007 with a proton beam energy of 920 GeV. The simulated detector response was processed and reconstructed using exactly the same software chain and the same procedures as being used for real data. The results of the processing were saved in ROOT [54] files in a form of ZEUS Common NTuples [37,38], a format that can be easily used for physics analysis without any ZEUS-specific software.

5 Event selection

The selection of events for the neural network training is motivated by the selection procedure applied in the previous ZEUS analyses [55–60]. Even though the presented analysis performed on the Monte Carlo simulated events, the selection cuts are choosen and applied as if the analysis is performed on real data for the purpose of being as close as possible to the measurements. The general motivation for these cuts

is the same as in many analyses performed by the ZEUS collaboration: to define unambiguously the phase space of the measurement, ensure low fraction of background events, and a reasonable description of the detector acceptance by Monte Carlo simulations.

5.1 Phase space selection

The phase space for the training of the neural networks in this analysis is selected as $100\,\mathrm{GeV}^2 < Q^2 < 20480\,\mathrm{GeV}^2$, being close to the phase space of the physics analysis in Ref. [59].

In the phase-space region at low x and very low inelasticity y, the QED predictions from the Monte Carlo simulations are not reliable because of a limit of higher orders in the calculations [61]. To avoid these phase-space regions, the events are required to have $y_{JB} \cdot (1 - x_{DA})^2 > 0.004$ and $y_{JB} > 0.04$ [61]. To ensure optimal electron identification and electron energy resolution, similar to the previous physics analyses, a kinematic cut $0.2 < y_{EL} < 0.7$ is used.

5.2 Event selection

The deep inelastic scattering events of interest are those characterised by the presence of a scattered electron in the final state and a significant deposit of energy in the calorimeter from the hadronic final state. The scattered electrons are registered in the detector as localised energy depositions primarily in the electromagnetic part of the calorimeter, with little energy flow into the hadronic part of the calorimeter. On the other hand, hadronic showers propagate in the detector much more extensively, both transversely and longitudinally.

In addition to the DIS process, there are also background processes which leave similar signatures in the detector as those described above. Therefore, the correct and efficient identification of the scattered lepton is crucial for the selection of the NC DIS events. For this analysis, the SINISTRA algorithm is used to identify lepton candidates [11]. Based on the information from the detector and the results of the SINISTRA algorithm, the following selection criteria are applied to select the events for the further analysis:

- **Detector status:** It is required that for all the events the detector was fully functional.
- **Electron energy:** At least one electron candidate with energy greater than $10\,\mathrm{GeV}$ [59] is identified in the event.
- **Electron identification probability:** The SINISTRA [11] probability of a lepton candidate being the DIS lepton was required to be greater than 90%. If several lepton candidates satisfy this condition, the candidate with the highest probability is used. In addition to this, there must be no problems reported by the SINISTRA algorithm.

- **Electron isolation:** To assist in removing events where the energy deposits from the hadron system overlap with those of the scattered lepton, the fraction of the energy not associated to the lepton is required to be less than 10% over the total energy deposited within a cone around the lepton candidate. The cone is defined with a radius of 0.7 units in the pseudorapidity-azimuth plane around the lepton momentum direction [59].
- **Electron track matching:** The tracking system covers the region of polar angles restricted to $0.3 < \theta < 2.85$ rad. Electromagnetic clusters within that region that have no matching track are most likely photons. If the lepton candidate is within this region, the presence of a matched track is required. This track must have a distance of closest approach between the track extrapolation point at the front surface of the CAL and the cluster centre-of-gravity-position of less than 10 cm. The track energy must be greater than $3\,\mathrm{GeV}$ [59].
- **Electron position:** To minimise the impact of imperfect simulation of some detector regions, additional requirements on the position of the electromagnetic shower are imposed. The events in which the lepton is found in the following regions (x, y and z being the cartesian coordinates in the ZEUS detector) are rejected [62]:
  - $\sqrt{x^2 + y^2} < 18$ cm, regions close the beam pipe
  - $z < -148$ cm and $y > 90$ cm and $-14 < x < 12$ cm, a part of the RCAL where the depth was reduced due to the cooling pipe for the solenoid (chimney),
  - $-104 < z < -98.5$ cm or $164 < z < 174$ cm, regions in-between calorimeter sections (super-crack).
- **Primary vertex position:** It was required that the reconstructed primary vertex position is close to the central region of the detector, applying the selection $-28.5 < Z_{\mathrm{vtx}} < 26.7$ cm [59].
- **Energy-longitudinal momentum balance:** To suppress photoproduction and beam-gas interaction background events and imperfect Monte Carlo simulations of those, restrictions are put on the energy-longitudinal momentum balance. This quantity is defined as:

$$\delta = \delta_l + \delta_{\mathcal{H}} = (E_{l'} - P_{z,l'}) + (E_{\mathcal{H}} - P_{z,\mathcal{H}}) = \sum_i (E_i - P_{z,i}), \tag{35}$$

where the final summation index runs over all energy deposits in the detector. In this analysis, we apply a condition of $38 < \delta < 65\,\mathrm{GeV}$ [59].
- **Missing transverse energy:** To remove beam-related background and cosmic-ray events, a cut on the missing energy is imposed. $P_{T,\mathrm{miss}}/\sqrt{E_T} < 2.5\,\mathrm{GeV}^{1/2}$, where
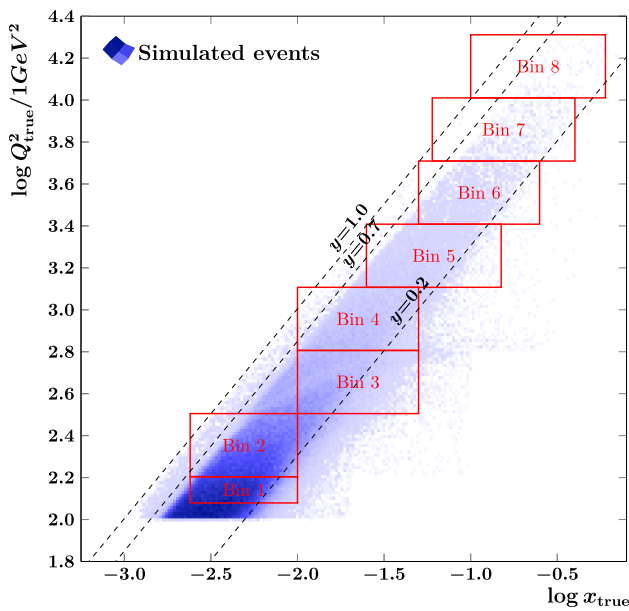
Fig. 3 Distribution of events from the training set in $(x_{\text{true}}, Q^2_{\text{true}})$ plane and the boundaries of the analysis bins from Table 3

E_T is the total transverse energy in the CAL and $P_{T,\text{miss}}$ is the missing transverse momentum, the transverse component of the vector sum of the hadronic final state and scattered electron momenta.

6 Training the DNN models

We use the neural network models defined in Sect. 3.3 and consider them as optimisation problem in terms of Eq. (29), i.e., we minimise the loss function across the selected training set and satisfy sparse regularity conditions. Every neural network making the ensemble model for x and Q^2 are trained simultaneously. The optimal regularity condition depends on the selection of the training set, the events batch size, the initial learning rate, and the regularisation parameter. We aim to select these in a way that minimises overfitting in particular regions of the kinematic space while still maximising the mean accuracy of the model.

We select a set of events from the "MEPS" data sets for training, as described in Sect. 5, and define the true values of the x and Q^2 as described in Sect. 2.3. Figure 3 shows a distribution of selected events in the $(x_{\text{true}}, Q^2_{\text{true}})$ plane and the boundaries of the chosen analysis bins.

First, we train the network to reconstruct Q^2 by optimising Eq. (29) with an initial learning rate of $\mathcal{L} = 1.0 \times 10^{-5}$ and a regularisation parameter of $\mathcal{R} = 1.0 \times 10^{-5}$. We select a batch size of 10,000.

The reconstruction of x is more complex. We fix the regularisation parameter to $\mathcal{R} = 1.0 \times 10^{-5}$ and select optimal parameters for the learning rate $\mathcal{L}$ and batch size $\mathcal{B}$ experi-

Table 1 Resolution of $\log x$ reconstruction after 200 epochs of training with different values of initial learning rate $\mathcal{L}$ and batch size $\mathcal{B}$

$\mathcal{L}$	RMS of $\log x - \log x_{\text{true}}$		
	$\mathcal{B}$ 10,000	$\mathcal{B}$ 50,000	$\mathcal{B}$ 100,000
1.0×10^{-7}	0.1507	0.1523	0.1598
5.0×10^{-7}	0.1568	0.1575	0.1575
1.0×10^{-6}	0.1504	0.1585	0.1547
5.0×10^{-6}	0.2384	0.2284	0.1829
1.0×10^{-5}	0.2182	2.5972	2.2767
1.5×10^{-5}	0.2122	1.7751	1.6028
2.0×10^{-5}	3.0258	0.2607	0.2852
5.0×10^{-5}	3.6962	2.4858	3.1231

Table 2 Resolution of $\log x$ reconstruction after 200 epochs of training with different values of regularisation parameter $\mathcal{R}$

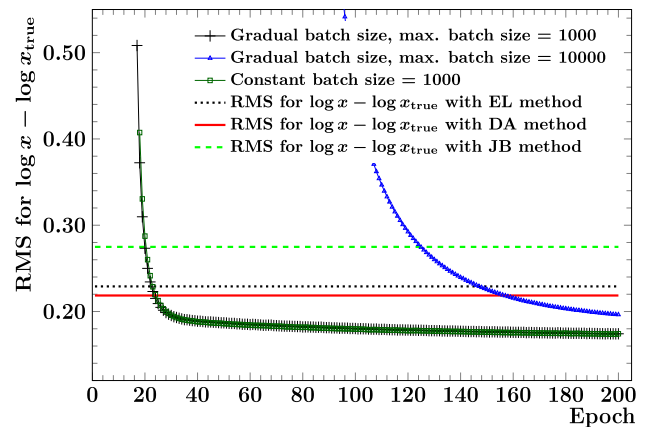| $\mathcal{R}$	RMS of $\log x - \log x_{\text{true}}$
1.0×10^{-6}	0.1493
1.0×10^{-5}	0.1494
1.0×10^{-4}	0.1484

Fig. 4 Training history for x reconstruction model using different training parameters but the same Q^2 reconstruction model obtained with $\mathcal{L} = 1.0 \times 10^{-5}$, $\mathcal{R} = 1.0 \times 10^{-6}$ and $\mathcal{B} = 10,000$. In each of the cases, the initial learning rate was set to $\mathcal{L} = 1.0 \times 10^{-5}$ and the regularisation parameter to $\mathcal{R} = 1.0 \times 10^{-6}$

mentally by varying the learning rate against the batch size. The appropriate selection of these parameters ensures fast convergence of the stochastic gradient method by balancing noise in the gradients with the stability of the algorithm. For each set of parameters, ten attempts are made and the best result in terms of the mean square error of the x reconstruction model over the training set is chosen. The results are listed in Table 1. The smaller learning rate assures a higher stability of the results than a larger learning rate. It prolongs the training process but avoids a poor convergence of the learn-

Table 3 Kinematic bins in x and Q^2, see also Fig. 3

Bin	Q^2 (GeV^2)	x
1	120–160	0.0024–0.010
2	160–320	0.0024–0.010
3	320–640	0.01–0.05
4	640–1280	0.01–0.05
5	1280–2560	0.025–0.150
6	2560–5120	0.05–0.25
7	5120–10240	0.06–0.40
8	10240–20480	0.10–0.60

ing process as observed for larger learning rates. The larger batch size does not offer any advantages in our analysis as shown in Table 1. We, therefore, select an initial learning rate of 10^{-5} with a minimal batch size. To choose the regularisation parameter close to the optimal one, we vary its value with constant batch size of 10,000 and initial learning rate of 10^{-5} and again observe the mean square error of the x reconstruction model over the training set. For each set of parameters, ten attempts are made and the best result is chosen. The results are presented in Table 2. Accordingly, we choose a regularisation parameter of 10^{-6}. Using this regularisation, the neural network models for both x and Q^2 are defined by weight parameters, of which 50% are effectively zero, or less than 10^{-8} Following the suggestions in Ref. [63], we start with a small batch size, and increase it in initial training epochs. We test this approach by comparing the mean square error of the x reconstruction model over the training set over the first 200 epochs of training over three different training regimes. The results are summarised in Fig. 4 and imply to use a gradually increasing batch size up to a maximum batch size of 1000.

7 Results

We evaluate the performance of our approach for the reconstruction of DIS kinematics by applying it to detailed Monte Carlo simulations from the ZEUS experiment and by comparing our results to the results from the electron, double-angle, and Jacques-Blondel reconstruction methods. For the comparison, we use various combinations of statistically independent data sets, one for the training, and another for the evaluation. In our systematic studies, we have found no signs of overtraining and also no indication that the results depend on the selected Monte Carlo simulations. For the results presented in this section, we use the "MEPS" data set for the training and the "CDM" data set for evaluation. The main quantities of the comparison are the resolutions of the reconstructed variables $\log Q^2/1\,\mathrm{GeV}^2$ and $\log x$ as measured in
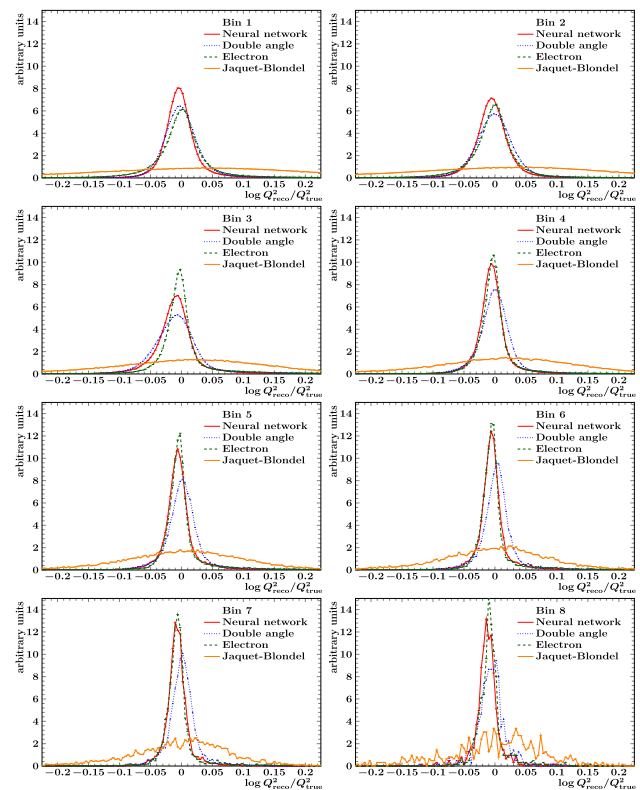

Fig. 5 Distributions of $\log Q^2/1\,\mathrm{GeV}^2 - \log Q_{\mathrm{true}}^2/1\,\mathrm{GeV}^2$ for various reconstruction methods in individual analysis bins. For better visibility, the data points for each reconstruction method are connected with straight lines

selected $x - Q^2$ regions (bins). The resolutions are defined as $\sqrt{\sum_i^N \left(\log Q_i^2/1\,\mathrm{GeV}^2 - \log Q_{i,\mathrm{true}}^2/1\,\mathrm{GeV}^2\right)^2 / N}$ and $\sqrt{\sum_i^N \left(\log x_i - \log x_{i,\mathrm{true}}\right)^2 / N}$, where N stands for the number events in the corresponding bin. The boundaries of the bins are given in Table 3 and are chosen to be close to the bins used in ZEUS DIS analyses, e.g. in Ref. [57].

The distributions of the $\log Q^2/1\,\mathrm{GeV}^2 - \log Q_{\mathrm{true}}^2/1\,\mathrm{GeV}^2$ and $\log x - \log x_{\mathrm{true}}$ quantities are given in the Figs. 5 and 6, respectively. The numerical values for the resolution are summarised for all the bins and methods in Table 4. The DNN optimisation procedure minimises the generalisation error described in Eq. (25) plus the regularisation penalty, so distributions in Figs. 5 and 6 do not necessarily peak at zero. In addition to that, Figs. 7 and 8 show the two dimensional distributions of events in $\log Q^2/1\,\mathrm{GeV}^2$ vs. $\log Q_{\mathrm{true}}^2/1\,\mathrm{GeV}^2$ and $\log x$ vs. $\log x_{\mathrm{true}}$ planes.

The comparison of the DNN-based approach with the classical methods demonstrates that the DNN-based approach is well suited for the reconstruction of DIS kinematics. Specifically, for most of the bins, our approach provides the best resolution as measured by the standard deviation of the logarithmic differences of true and reconstructed variables. The
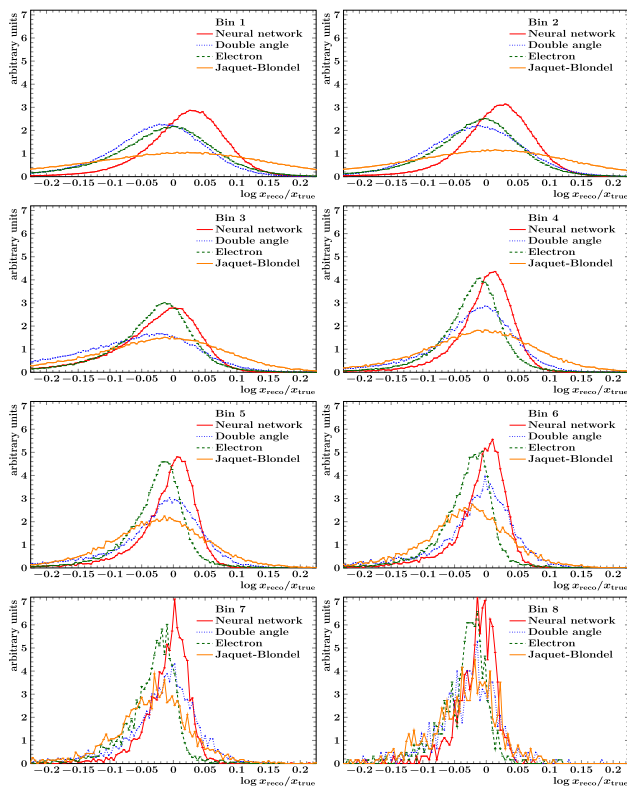
Fig. 6 Distributions of $\log x - \log x_{\text{true}}$ for various reconstruction methods in individual analysis bins. For better visibility, the data points for each reconstruction method are connected with straight lines

better performance of the DNN-based approach in most of the bins is a consequence of using additional available information about the final state. In this respect, the DNN-based approach is similar to averaging of the values provided by the classical methods with some weights or to alternative approaches for the same task, e.g. see kinematic fitting in Ref. [64]. However, in addition to a better resolution, the reconstruction with the DNN has an important advantage over the classical methods or any simple combination of them. It allows to combine the methods without the intrinsic biases of each method and enables an extension of the model with additional physics observables in a robust way.

The resolution is improved by the DNN-based approach in two ways. For the first couple of bins, the main improvement is caused by the more precise estimation of the reconstructed observables. This is clearly seen in Figs. 5 and 6. For the bins with higher Q^2 and x the main improvement is due to the rejection of outliers, which can be seen in Figs. 7 and 8. The bins with higher Q^2 and x also demonstrate another, very specific advantage on the DNN approach. Due to the low number of training events in the high Q^2 and x region, it would not be possible to train a DNN model or combine the Q^2 and x observables with other methods using information from this region only. However, the DNN training process benefits from the constraints from the higher number of events available elsewhere in the kinematical space and delivers models that perform well even in the bins with highest Q^2 and x.

Table 4 Resolution of the reconstructed kinematic variables in bins of x and Q^2. The resolution for x and Q^2 is defined as the RMS of the distributions $\log(x) - \log(x_{\text{true}})$ and $\log(Q^2) - \log(Q^2_{\text{true}})$ respectively

| Bin	Events	Resolution of $\log x$		Resolution of $\log Q^2 / 1\,\text{GeV}^2$	
1	301780	**NN: 0.070**	EL: 0.083	**NN: 0.035**	EL: 0.035
		JB: 0.180	DA: 0.103	JB: 0.203	DA: 0.062
2	350530	**NN: 0.069**	EL: 0.082	**NN: 0.040**	EL: 0.043
		JB: 0.167	DA: 0.096	JB: 0.192	DA: 0.064
3	138456	**NN: 0.098**	EL: 0.130	NN: 0.055	**EL: 0.053**
		JB: 0.138	DA: 0.100	JB: 0.150	DA: 0.077
4	74844	**NN: 0.067**	EL: 0.084	**NN: 0.044**	EL: 0.046
		JB: 0.117	DA: 0.077	JB: 0.138	DA: 0.063
5	31043	**NN: 0.064**	EL: 0.091	**NN: 0.036**	EL: 0.041
		JB: 0.102	DA: 0.073	JB: 0.117	DA: 0.053
6	11475	**NN: 0.053**	EL: 0.079	**NN: 0.033**	EL: 0.036
		JB: 0.083	DA: 0.061	JB: 0.100	DA: 0.045
7	3454	**NN: 0.050**	EL: 0.069	**NN: 0.036**	EL: 0.038
		JB: 0.074	DA: 0.055	JB: 0.093	DA: 0.042
8	624	**NN: 0.036**	EL: 0.055	**NN: 0.033**	EL: 0.037
		JB: 0.067	DA: 0.045	JB: 0.095	DA: 0.041

Minimal value shown in bold
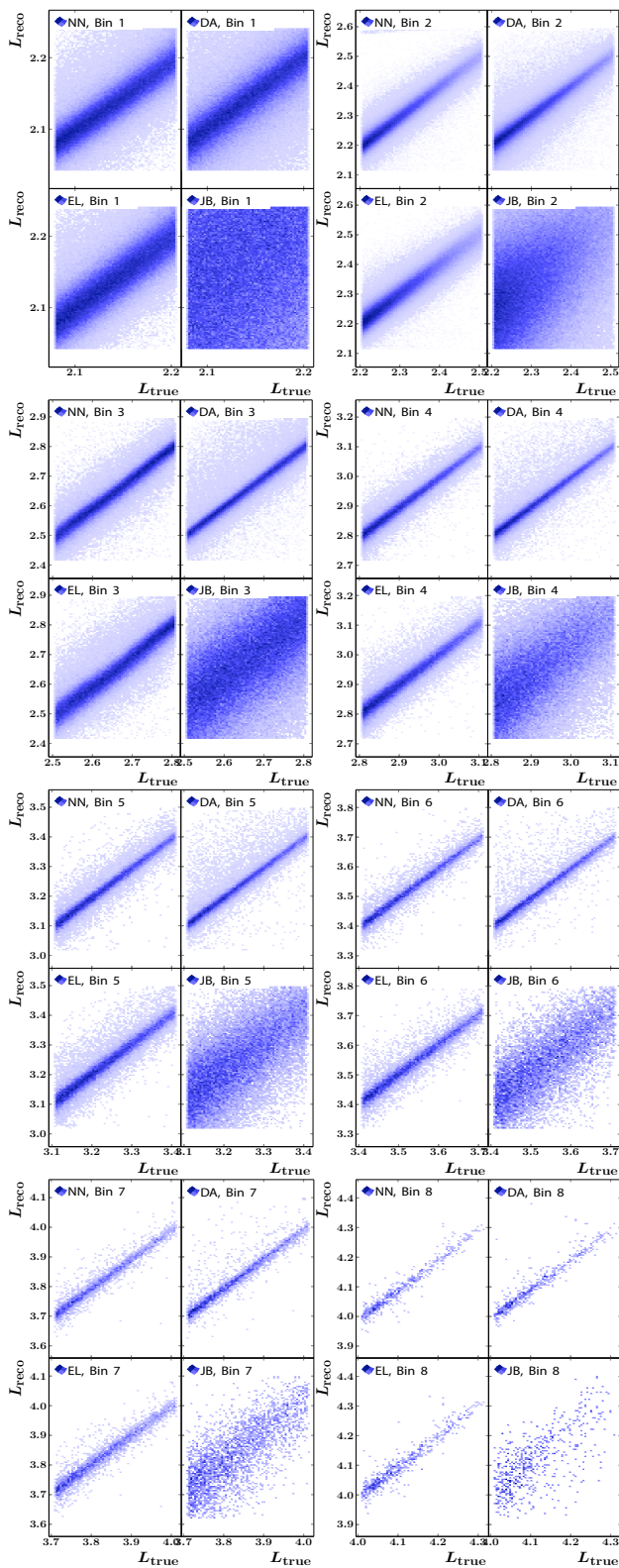
Fig. 7 Distribution of events in $L_{reco} = \log Q^2/1\,\text{GeV}^2$ versus $L_{true} = \log Q^2_{true}/1\,\text{GeV}^2$ plane for various reconstruction methods in individual analysis bins
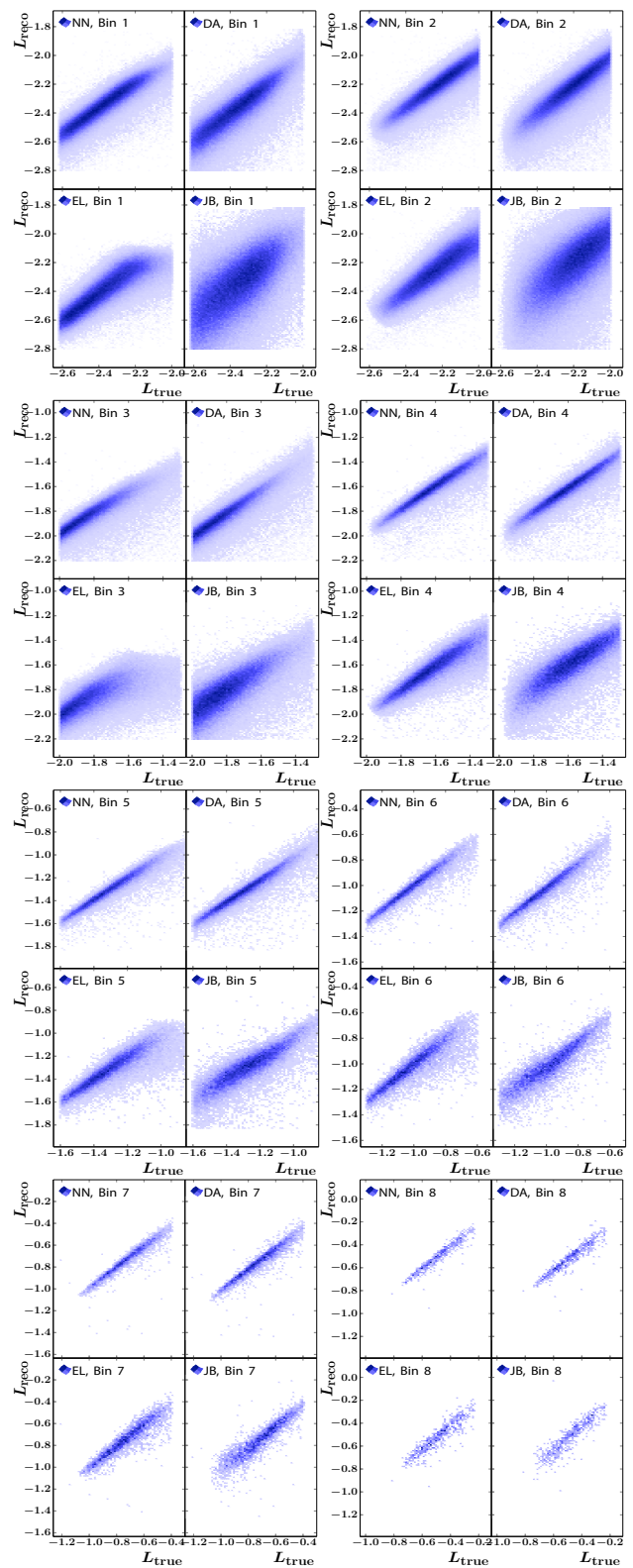
Fig. 8 Distribution of events in $L_{reco} = \log x$ versus $L_{true} = \log x_{true}$ plane for various reconstruction methods in individual analysis bins

8 Conclusions

We have presented the use of DNN to reconstruct the kinematic observables Q^2 and x in the study of neutral current DIS events at the ZEUS experiment at HERA. The DNN models are specially designed to be effective in their universal approximation capability, robust in the sense that increasing the depth of the networks will necessarily reduce empirical error, and computationally efficient with a structure that avoids "vanishing" gradients arising in the backpropagation algorithm.

Compared to the classical reconstruction methods, the DNN-based approach enables significant improvements in the resolution of Q^2 and x. At the same time, it is evident that the usage of the DNN approach allows to match easily any definition of Q^2 and x at the true level that is preferred for a given physics analysis.

The large samples of simulated data required for the training of the DNN can be generated rapidly at modern data centres. Also, DNN allow to effectively extract information from large data sets. This suggests that our new approach for the reconstruction of DIS kinematics can serve as a rigorous method to combine and outperform the classical reconstruction methods at ongoing or upcoming DIS experiments. We will extend the approach beyond inclusive DIS measurements and will study next the use of DNN for the reconstruction of event kinematics in semi-inclusive and exclusive DIS measurements.

Appendix: Software used in the analysis

The `ROOT` package [54] of version 6.22 was used to read the ZEUS data, provided by the data preservation at Max-Planck for Physics [65], analyze it and prepare plain text (or ROOT files) with selected information to be used with the ML tools. The selected information from the plain text (or ROOT) files was piped using the `pandas` package [66] into `Keras` [67] interface to `tensorflow` [68] 2.3.0 library to train the ML models. The packages `Eigen` [69], `frugally-deep` [70], `JSON for Modern C++` [71] and `FunctionalPlus` [72] were used for execution of the trained models after these were converted into `frugally-deep` model format [70] to be used with C++ application. The dependencies for the `tensorflow` were supplied from the `PyPi` repository. The training was performed using libraries for computing on GPUs from the `CUDA` [73] framework of version 10.1. We are grateful to MPCDF[1] for the ability to compile and execute the codes on the HPC cluster "Cobra" in [74]. The operation system used was Linux on x86_64 architecture using `gcc` [75] of version 7.3 and `python` [76] of version 3.6.8.

The figures with the final results were produced with the `PGFPlots` [77] package.

References

1. B.H. Wiik et al., HERA: A Proposal for a Large Electron Proton Colliding Beam Facility at DESY (1981). https://old.inspirehep.net/record/19436/files/Fulltext.pdf
2. DUNE Coll., B. Abi et al., Volume I. Introduction to DUNE. JINST **15**, T08008 (2020). https://doi.org/10.1088/1748-0221/15/08/T08008. arXiv:2002.0296
3. F. Gautheron et al., COMPASS-II Proposal (2010). https://cds.cern.ch/record/1265628/files/SPSC-P-340.pdf
4. J. Arrington et al., Physics with CEBAF at 12 GeV and Future Opportunities (2021). arXiv:2112.0006
5. E.C. Aschenauer, R.S. Thorne, R. Yoshida, PDG Chapter 18: Structure Functions. PTEP (2022). https://doi.org/10.1093/ptep/ptac097
6. G.A. Voss, B.H. Wiik, The Electron proton collider HERA. Ann. Rev. Nucl. Part. Sci. **44**, 413 (1994). https://doi.org/10.1146/annurev.ns.44.120194.002213
7. H1 and ZEUS Colls., H. Abramowicz et al., Combination of measurements of inclusive deep inelastic $e^{\pm}p$ scattering cross sections and QCD analysis of HERA data. Eur. Phys. J. C **75**, 580 (2015). https://doi.org/10.1140/epjc/s10052-015-3710-4. arXiv:1506.0604
8. U. Bassler, G. Bernardi, On the kinematic reconstruction of deep inelastic scattering at HERA: the Sigma method. Nucl.

Instrum. Methods A **361**, 197 (1995). https://doi.org/10.1016/0168-9002(95)00173-5. arXiv:hep-ex/9412004

9. R. Abdul Khaleket et al., Science requirements and detector concepts for the electron-ion collider: EIC Yellow Report (2021). arXiv:2103.0541

10. U. Holm, The ZEUS detector: Status report 1993 (1993). https://doi.org/10.3204/PUBDB-2017-12635

11. H. Abramowicz, A. Caldwell, R. Sinkus, Neural network based electron identification in the ZEUS calorimeter. Nucl. Instrum. Methods A **365**, 508 (1995). https://doi.org/10.1016/0168-9002(95)00612-5. arXiv:hep-ex/9505004

12. A. Accardi et al., Electron ion collider: the next QCD Frontier. Eur. Phys. J. A **52**, 268 (2016). https://doi.org/10.1140/epja/i2016-16268-9. arXiv:1212.1701

13. J. Liu et al., Deep-Learning-Based Kinematic Reconstruction for DUNE (2020). arXiv:2012.0618

14. H1, I. Abt et al., The H1 detector at HERA. Nucl. Instrum. Methods A **386**, 310 (1997) https://doi.org/10.1016/S0168-9002(96)00893-5

15. A. Kwiatkowski, H. Spiesberger, H.J. Mohring, Heracles: an event generator for *ep* interactions at HERA energies including radiative processes: version 1.0. Comput. Phys. Commun. **69**, 155 (1992) https://doi.org/10.1016/0010-4655(92)90136-M

16. G. Altarelli, G. Parisi, Asymptotic freedom in parton language. Nucl. Phys. B **126**, 298 (1977). https://doi.org/10.1016/0550-3213(77)90384-4

17. T. Sjöstrand, S. Mrenna, P.Z. Skands, PYTHIA 6.4 physics and manual. JHEP **05**, 026 (2006). https://doi.org/10.1088/1126-6708/2006/05/026. arXiv:hep-ph/0603175

18. T. Sjöstrand et al., An introduction to PYTHIA 8.2. Comput. Phys. Commun. **191**, 159 (2015). https://doi.org/10.1016/j.cpc.2015.01.024. arXiv:1410.3012

19. Sherpa Coll., E. Bothmann et al., Event generation with SHERPA 2.2. SciPost Phys. **7**, 034 (2019). https://doi.org/10.21468/SciPostPhys.7.3.034. arXiv:1905.0912

20. W. Kilian, T. Ohl, J. Reuter, WHIZARD: simulating multi-particle processes at LHC and ILC. Eur. Phys. J. C **71**, 1742 (2011). https://doi.org/10.1140/epjc/s10052-011-1742-y. arXiv:0708.4233

21. J. Bellm et al., Herwig 7.0/Herwig++ 3.0 release note. Eur. Phys. J. C **76**, 196 (2016). https://doi.org/10.1140/epjc/s10052-016-4018-8. arXiv:1512.0117

22. G. Ingelman, A. Edin, J. Rathsman, LEPTO 6.5: A Monte Carlo generator for deep inelastic lepton–nucleon scattering. Comput. Phys. Commun. **101**, 108 (1997). https://doi.org/10.1016/S0010-4655(96)00157-9. arXiv:hep-ph/9605286

23. L. Lönnblad, ARIADNE version 4: a program for simulation of QCD cascades implementing the color dipole model. Comput. Phys. Commun. **71**, 15 (1992). https://doi.org/10.1016/0010-4655(92)90068-A

24. H. Jung et al., The CCFM Monte Carlo generator CASCADE version 2.2.03. Eur. Phys. J. C **70**, 1237 (2010). https://doi.org/10.1140/epjc/s10052-010-1507-z. arXiv:1008.0152

25. H. Jung, Hard diffractive scattering in high-energy *ep* collisions and the Monte Carlo generator RAPGAP. Comput. Phys. Commun. **86**, 147 (1995). https://doi.org/10.1016/0010-4655(94)00150-Z

26. F. Cascioli, P. Maierhofer, S. Pozzorini, Scattering amplitudes with OpenLoops. Phys. Rev. Lett. **108**, 111601 (2012). https://doi.org/10.1103/PhysRevLett.108.111601. arXiv:1111.5206

27. Z. Bern et al., The BlackHat library for one-loop amplitudes. J. Phys. Conf. Ser. **523**, 012051 (2014). https://doi.org/10.1088/1742-6596/523/1/012051. arXiv:1310.2808

28. J. Alwall et al., MadGraph 5: going beyond. JHEP **06**, 128 (2011). https://doi.org/10.1007/JHEP06(2011)128. arXiv:1106.0522

29. R. Brun et al., GEANT3. https://cds.cern.ch/record/1119728/files/CERN-DD-EE-84-1.pdf (1987)

30. GEANT4 Coll., S. Agostinelli et al., GEANT4–a simulation toolkit. Nucl. Instrum. Methods A **506**, 250 (2003). https://doi.org/10.1016/S0168-9002(03)01368-8

31. S. Bentvelsen, J. Engelen, P. Kooijman, Reconstruction of (x, Q^2) and extraction of structure functions in neutral current scattering at HERA, Workshop on Physics at HERA Hamburg, Germany, October 29-30, 1991, pp. 23–42 (1992)

32. J. Currie et al, NNLO QCD corrections to jet production in deep inelastic scattering. JHEP **07**, 018 (2017). (Erratum: JHEP 12, 042 (2020)). https://doi.org/10.1007/JHEP07(2017)018. arXiv:1703.0597

33. H1, Andreev, V. and others, Measurement of lepton-jet correlation in deep-inelastic scattering with the H1 detector using machine learning for unfolding. Phys. Rev. Lett. **128**, 132002 (2022). https://doi.org/10.1103/PhysRevLett.128.132002. arXiv:2108.1237

34. A. Arbuzov el al., Hector 1.00: a program for the calculation of QED, QCD and electroweak corrections to *ep* and $lepton^{\pm}N$ deep inelastic neutral and charged current scattering. Comput. Phys. Commun. **94**, 128 (1996). https://doi.org/10.1016/0010-4655(96)00005-7. arXiv:hep-ph/9511434

35. T. Liu et al., A new approach to semi-inclusive deep-inelastic scattering with QED and QCD factorization. JHEP **11**, 157 (2021). https://doi.org/10.1007/JHEP11(2021)157. arXiv:2108.1337

36. ZEUS, H. Abramowicz et al., Inclusive-jet cross sections in NC DIS at HERA and a comparison of the kT, anti-kT and SIScone jet algorithms. Phys. Lett. B **691**, 127 (2010). https://doi.org/10.1016/j.physletb.2010.06.015. arXiv:1003.2923

37. ZEUS Coll., J. Malka, K. Wichmann, The ZEUS data preservation project. J. Phys. Conf. Ser. **396**, 022033 (2012). https://doi.org/10.1088/1742-6596/396/2/022033

38. ZEUS Coll., A. Verbytskyi, The ZEUS long term data preservation project. PoS **DIS2016**, 264 (2016). https://doi.org/10.22323/1.265.0264. arXiv:1607.0189

39. K. He et al., Deep residual learning for image recognition (2015). arXiv:1512.0338

40. A. Farhat, Deep multiscale learning and applications to nuclear physics, Ph.D. thesis, Old Dominion University, in preparation

41. K. Hornik, M. Stinchcombe, H. White, Multilayer feedforward networks are universal approximators. Neural Netw. **2**, 359 (1989)

42. M. Leshno, V.Y. Lin, A. Pinkus, S. Schocken, Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. Neural Netw. **6**, 861 (1993)

43. E. Candes, J. Romberg, T. Tao, Robust uncertainty principles: exact signal frequency information. IEEE Trans. Inf. Theory **52**, 489 (2006). https://doi.org/10.1109/TIT.2005.862083

44. H. Zhang, Y. Xu, J. Zhang, Reproducing kernel Banach spaces for machine learning. J. Mach. Learn. Res. 2741 (2009)

45. D.P. Bertsekas, J. Tsitsiklis, Gradient convergence in gradient methods with errors. SIAM J. Optim. **10**, 627 (2000)

46. I. Sutskever et al., On the importance of initialization and momentum in deep learning. ICML'13: Proceedings of the 30th International Conference on International Conference on Machine Learning **28**, pp. 1139–1147 (2013)

47. Y. Xu, H. Zhang, Convergence of deep relu networks (2021). arXiv:2107.1253

48. Y. Xu, H. Zhang, Convergence of deep convolutional neural networks. Neural Netw. **153**, 553 (2022)

49. A. Krizhevsky et al., ImageNet classification with deep convolutional neural networks, in *Advances in Neural Information Processing Systems* ed. by F. Pereira, C.J. C. Burges, L. Bottou, K.Q. Weinberger, vol. 25. (Curran Associates, Inc., 2012), p. 1097–1105

50. G.A. Schuler, H. Spiesberger, DJANGO: the Interface for the event generators HERACLES and LEPTO, Workshop on Physics at HERA (1991)

51. CTEQ Coll., H.L. Lai et al., Global QCD analysis of parton structure of the nucleon: CTEQ5 parton distributions. Eur.

Phys. J. C **12**, 375 (2000). https://doi.org/10.1007/s100529900196. arXiv:hep-ph/9903282

52. M. Bengtsson, T. Sjostrand, Parton showers in leptoproduction events. Z. Phys. C **37**, 465 (1988). https://doi.org/10.1007/BF01578142

53. G. Gustafson, U. Pettersson, Dipole formulation of QCD cascades. Nucl. Phys. B **306**, 746 (1988). https://doi.org/10.1016/0550-3213(88)90441-5

54. I. Antcheva et al., ROOT: A C++ framework for petabyte data storage, statistical analysis and visualization. Comput. Phys. Commun. **182**, 1384 (2011). https://doi.org/10.1016/j.cpc.2011.02.008

55. ZEUS Coll., J. Breitweg et al., Measurement of high Q^2 neutral current $e^+ p$ deep inelastic scattering cross-sections at HERA. Eur. Phys. J. C **11**, 427 (1999). https://doi.org/10.1007/s100520050645. arXiv:hep-ex/9905032

56. ZEUS Coll., S. Chekanov et al., Jet-radius dependence of inclusive-jet cross-sections in deep inelastic scattering at HERA. Phys. Lett. B **649**, 12 (2007). https://doi.org/10.1016/j.physletb.2007.03.039. arXiv:hep-ex/0701039

57. ZEUS Coll., S. Chekanov et al., Event shapes in deep inelastic scattering at HERA. Nucl. Phys. B **767**, 1 (2007). https://doi.org/10.1016/j.nuclphysb.2006.05.016. arXiv:hep-ex/0604032

58. ZEUS Coll., S. Chekanov et al., Inclusive-jet and dijet cross-sections in deep inelastic scattering at HERA. Nucl. Phys. B **765**, 1 (2007). https://doi.org/10.1016/j.nuclphysb.2006.09.018. arXiv:hep-ex/0608048

59. ZEUS Coll., H. Abramowicz et al., Inclusive-jet cross sections in NC DIS at HERA and a comparison of the kT, anti-kT and SIScone jet algorithms. Phys. Lett. B **691**, 127 (2010). https://doi.org/10.1016/j.physletb.2010.06.015. arXiv:1003.2923

60. ZEUS Coll., H. Abramowicz et al., Inclusive dijet cross sections in neutral current deep inelastic scattering at HERA. Eur. Phys. J. **C70**, 965 (2010). https://doi.org/10.1140/epjc/s10052-010-1504-2, arXiv:1010.6167

61. D. Lontkovskyi, Measurement of Jet Production with the ZEUS Detector, Ph.D. thesis, Hamburg U., 2015. http://ediss.sub.uni-hamburg.de/volltexte/2016/7801/

62. H. Perrey, Jets at low Q^2 at HERA and radiation damage studies for silicon sensors for the XFEL, Ph.D. thesis, Hamburg U (2011). http://www-library.desy.de/preparch/desy/thesis/desy-thesis-11-021.pdf

63. S. Smith, P. Kindermans, Q. Le, Don't decay the learning rate, increase the batch size (2017). arXiv:1711.0048

64. R. Aggarwal, Measurement of high x neutral current ep cross sections and extractions of $x F^3$ structure function using ZEUS detector at HERA, Ph.D. thesis, Panjab University (2012). http://hdl.handle.net/10603/80282

65. DPHEP Coll., S. Amerio et al., Status Report of the DPHEP Collaboration: a global effort for sustainable data preservation in high energy physics. zenodo (2015). https://doi.org/10.5281/zenodo.46158. arXiv:1512.0201

66. W. McKinney, Data Structures for Statistical Computing in Python, in *Proceedings of the 9th Python in Science Conference*, ed. by S. van der Walt, J. Millman (2010), p. 56–61. https://doi.org/10.25080/Majora-92bf1922-00a

67. F. Chollet et al., Keras. https://keras.io, (2015)

68. M. Abadi et al., TensorFlow: large-scale machine learning on heterogeneous systems. http://tensorflow.org/ (2015)

69. G. Guennebaud, B. Jacob et al., Eigen v3. http://eigen.tuxfamily.org, (2021)

70. T. Hermann, frugally-deep. https://github.com/Dobiasd/frugally-deep (2021)

71. N. Lohmann, JSON for Modern C++. https://github.com/nlohmann/json (2021)

72. T. Hermann, FunctionalPlus. https://github.com/Dobiasd/FunctionalPlus (2021)

73. NVIDIA Corp., P. Vingelmann, F.H.P Fitzek, CUDA, release: 10.2.89. https://developer.nvidia.com/cuda-toolkit (2020)

74. Max-Planck Computing and Data Facility, About the HPC system Cobra. https://www.mpcdf.mpg.de/services/supercomputing/cobra (2021)

75. R.M. Stallman and GCC Developer Community, Using The Gnu Compiler Collection: A Gnu Manual For GCC Version 4.3.3. CreateSpace, Paramount, CA (2009)

76. Python Software Foundation, Python Language Reference, version 3. http://www.python.org (2019)

77. C. Feuersänger, PGFPlots. http://pgfplots.sourceforge.net/ (2021)