Analytica Chimica Acta 1271 (2023) 341433

Contents lists available at ScienceDirect

Analytica Chimica Acta

journal homepage: www.elsevier.com/locate/aca

Convolutional neural network framework for the automated analysis of transition metal X-ray photoelectron spectra

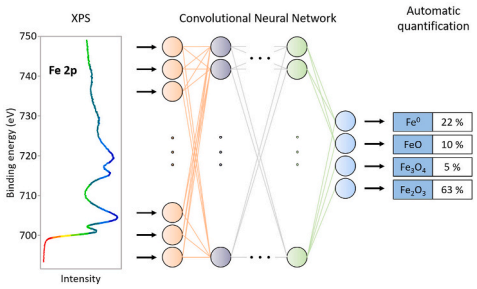
Lukas Pielsticker *, Rachel L. Nicholls, Serena DeBeer, Mark Greiner

Max Planck Institute for Chemical Energy Conversion, Stiftstr. 34-36, 45470, Muelheim an der Ruhr, Germany

HIGHLIGHTS

- XPS data are of increasing complexity, necessitating automated analysis approaches.
- Convolutional neural networks enable the acceleration of the analysis process.
- Uncertainty measure is used to identify non-standard chemical species in XP spectra.
- Useful for researchers investigating structure-function correlations using XPS.

GRAPHICAL ABSTRACT



ARTICLE INFO

Handling Editor: Xiu-Ping Yan

Keywords:
Electron spectroscopy
Transition metals
Supervised machine learning
Convolutional neural networks
Automated analysis

ABSTRACT

X-ray photoelectron spectroscopy is an indispensable technique for the quantitative determination of sample composition and electronic structure in diverse research fields. Quantitative analysis of the phases present in XP spectra is usually conducted manually by means of empirical peak fitting performed by trained spectroscopists. However, with recent advancements in the usability and reliability of XPS instruments, ever more (inexperienced) users are creating increasingly large data sets that are harder to analyze by hand. In order to aid users with the analysis of large XPS data sets, more automated, easy-to-use analysis techniques are needed. Here, we propose a supervised machine learning framework based on artificial convolutional neural networks. By training such networks on large numbers of artificially created XP spectra with known quantifications (i.e., for each spectrum, the concentration of each chemical species is known), we created universally applicable models for auto-quantification of transition-metal XPS data that are able to predict the sample composition from spectra within seconds. Upon evaluation against more traditional peak fitting methods, we showed that these neural networks achieve competitive quantification accuracy. The proposed framework is shown to be flexible enough to accommodate spectra containing multiple chemical elements and measured with different experimental parameters. The use of dropout variational inference for the determination of quantification uncertainty is illustrated.

* Corresponding author.
E-mail address: lukas.pielsticker@cec.mpg.de (L. Pielsticker).

https://doi.org/10.1016/j.aca.2023.341433
Received 18 January 2023; Received in revised form 15 May 2023; Accepted 26 May 2023
Available online 27 May 2023

1. Introduction

X-ray photoelectron spectroscopy (XPS) is one of the most widely used techniques to investigate surface chemistry and electronic band structures in various research fields including chemistry, materials science, and mechanical engineering. Due to the short electron attenuation lengths in typical sample materials, XPS can be used for surface-sensitive quantitative analysis of a sample surface. Typical probing depths are between 1 and 5 nm of the sample surface for Al Kα radiation. XPS is frequently employed to determine the sample composition and quantify chemical species in the near-surface region.

In XPS, quantification of a given chemical species i is typically performed by determining the integrated area A_i under the corresponding peak(s) and normalizing this area by relative sensitivity factors (RSF), which take into account the photoionization cross-section for the studied transition, to obtain a normalized intensity $I_i = A_i/\mathrm{RSF}$. Assuming a homogeneous distribution of species in the analyzed sample region, the atomic fraction X_i of each phase is its normalized intensity divided by the sum of normalized intensities of all other phases:

$$X_i = \frac{I_i}{\sum_j I_j}. \quad (1)$$

The problem with this approach lies in the precise determination of the areas A_i. This generally involves a non-trivial empirical fit of both the inelastic scattering background and the zero energy-loss line shapes of the peak regions. Since there is typically no closed analytical function for either background or line shapes, peak fitting necessarily involves practical assumptions and simplifications. Therefore, this approach is limited by many uncertainties, including the choice of fitting background, as well as the assignment and shape of peaks [1,2]. Especially for the complex spectra of transition metals, quantification can be complicated due to asymmetries in peak shapes, complex satellite structures, multiplet splitting, bulk and surface plasmon losses, and overlapping line shapes [3]. Moreover, translational shifts due to charging and incorrect energy calibration can further complicate the accuracy of peak assignment [4]. Quantitative XPS can additionally be influenced by the measurement instrument, as both the binding energy position of each peak and the energy resolution depend on the hardware models (hemispherical analyzer, detector) and their settings. In total, these error sources can lead to error bars in the range of 10–20%, well above what is needed to accurately quantify species concentrations in XPS [5–9].

While a trained spectroscopist may be able to limit the quantification error by using best practices [1,7], the presence of noise in the data and changes in the energy resolution (e.g., when different spectra were measured with varying pass energies) often inhibit the transfer of a working fit model from one spectrum to another. Therefore, the fit model needs to be adjusted for each spectrum, which limits the applicability of the traditional analysis approach to large data sets. However, due to the advancements in X-ray sources (including the use of high-brilliance synchrotron radiation) and electron detectors, ever larger data sets are obtained in high-throughput experiments [10]. In turn, data processing that matches this high throughput is required to be able to keep up with these developments. Moreover, XPS instruments are becoming increasingly more reliable and automated, making it easier for novice users to perform XPS measurement. Oftentimes, these new users lack the experience of XPS experts and are therefore not well-suited to manually perform correct XPS analysis. But, even the experienced users make assumptions and mistakes in manual fitting. Thus, a more automated way of quantitative analysis would benefit the whole community of XPS users. Additionally, there has been a new push towards building reference databases of X-ray spectroscopy data [11]. Automatic labeling of these reference spectra with the correct quantification of chemical species would significantly enhance the usability of these databases. In recognition of the significant challenges outlined above, several attempts at automation of XPS peak fitting methods have been undertaken in recent years [10,12–14], mostly focused on automating the development of curve fitting methods. However, while certainly useful, none of these approaches has been able to fully automate the quantitative analysis of a large number of XP spectra from different elements.

In order to find a new approach to the automation of XPS analysis, we developed a method of automatically quantifying the chemical components in transition metal (TM) XPS spectra through the use of a Convolutional Neural Network (CNN). While originally developed for pattern recognition problems, such as computer vision and voice recognition, CNNs have been shown to aid data analysis in several spectroscopy applications, including Raman spectroscopy, Electron Energy Loss Spectroscopy (EELS) and X-ray Absorption Spectroscopy (XAS) [15–27]. For example, Chatzidakis et al. demonstrated the use of a CNN to quantify different Mn oxide species in EELS data [17]. For XPS, there have not been many studies using neural networks up to this point. In a recent study by Drera et al., it was demonstrated that CNNs can be used to quantify the stoichiometry of each element present in XP survey spectra in agreement with experimental results, within an error of 10% [18]. However, with this approach it was only possible to quantify the presence of different elements, but not to determine the identity and atomic percentages of different chemical states of these elements.

In this work, we present a framework for automatically determining the concentration of different chemical phases in transition metal XP spectra. In the absence of sufficiently large experimental data sets, artificial spectra were created based on well-characterized reference spectra, such that for each artificial spectrum the 'ground truth' stoichiometry, that is, the concentration of each chemical species, is known. Typical experimental artifacts such as binding energy shifts, peak broadening, varying signal-to-noise (S/N) ratio, as well as environmental influences such as the scattering of the photoelectrons in a gas phase, were simulated in order to produce randomized synthetic, yet realistic spectra. By training convolutional neural networks on large numbers of such artificially created XP spectra with known stoichiometries, we created universally applicable models for auto-quantification.

2. Materials and methods

2.1. Experimental

All XP spectra presented in this work were measured using a Phoibos NAP-150 hemispherical analyzer and a monochromatic Al Kα source from Specs GmbH. Metal foil samples were sputter cleaned using argon ions (with an energy of 2.5 kV, at a pressure of 1×10^{-5} mbar, and an emission current of 10 mA, for 30 min), to remove adventitious carbon and any surface oxides, allowing for the measurement of clean metallic references. Oxide reference spectra were obtained by heating and cooling oxide reference materials (Sigma-Aldrich) in O_2 and H_2 in the temperature range of 300 K and 1273 K. Detailed information about the procedure to obtain the different reference spectra can be found in the supplementary information (Table S1). During XPS measurements where gas was present, gas was leaked into the sample chamber using mass flow controllers. A throttle valve controlling the pumping cross section of a differential pumping stage was used to keep the chamber pressure constant during measurements. The pressure was measured using a diaphragm capacitance pressure gauge. All reference spectra that were used in this work are shown in the supporting information (Fig. S1).

2.2. Data set generation

In order to obtain sufficiently large data sets for the training of convolutional neural networks, XP spectra had to be calculated numerically. In the literature, there exist many methods for simulating XP spectra based on underlying theory [28,29]. However, for our work, existing methods were not able to generate spectra that took into

account all of the characteristics that we wanted to model. Specifically, these established methods could not take into account the presence of Auger spectra (from the same and from other elements) that can coincide with the core-level XPS regions, depending on the X-ray excitation energies that are used. Therefore, in this work, the simulated spectra are not based on theory, but rather on well-characterized spectra from reference materials.

In order to create an artificial mixed metal-oxide spectrum, reference core-level and Auger spectra of the pure metal and oxide phases were measured with high precision using lab-based XPS in vacuum as well as in near-ambient pressure (NAP-XPS). Detailed information about the procedure to obtain the reference spectra are shown in Table S1 in the supporting information. An artificial spectrum was created by linear combination of the measured metal and oxide reference spectra, with randomly selected stoichiometries. The stoichiometry ratios were scaled such that the overall atomic percentage of all species added up to 100%. The minimum contribution of each reference was set to 1%, in line with typical detection limits in XPS [30]. The spectrum resulting from the linear combination of the selected reference spectra was then shifted and distorted in several ways in order to emulate the variance in position on the binding energy axis, signal-to-noise ratio, and measurement resolution during real-world XPS experiments.

The experimental resolution during an XPS measurement is determined by a number of chemical as well as instrumental influences. Peak broadening can occur for a variety of reasons, including the choice of pass energy of the hemispherical electron analyzer during the experiment and differential charging of the surface. In order to reproduce these effects, the simulated spectrum was convolved with a Gaussian peak to simulated broadening. The full width at half maximum (FWHM) of the Gaussian peak was chosen from the interval [0.5 eV, 3.0 eV]. In addition to broadening, there is the frequent possibility of peak shifts by several eV during XPS measurement, which result from (differential) charging or changes to the analyzer-to-sample work function. These effects were taken into account by shifting each synthetic spectrum in the range of [−3 eV, 3 eV]. Moreover, changes in the signal-to-noise ratio were introduced by applying Poisson-distributed noise of a random intensity to each spectrum. The use of Poisson statistics to simulate noise is justified since XPS data is Poisson-distributed due to the single event nature of the detection of electrons at the detector [31]. Signal-to-noise (S/N) ratios in the range of [2,35] were simulated. Recently, the focus of some XPS experimenters has been shifted towards measurements in non-ultra-high vacuum conditions, such as in near-ambient pressure XPS (NAP-XPS) [32,33]. In such experiments, the resulting spectra are distorted because of inelastic scattering of the outgoing photoelectrons with the gas phase atoms [34,35]. In order to allow for accurate phase quantification from spectra taken during NAP-XPS experiments, the inelastic scattering by the gas phase was also simulated. This involves convolving the spectrum with the inelastic energy loss function of the gas-phase scattering medium. Here, loss functions of H_2, He, O_2, and N_2 were used to simulate the scattering in some gases that are frequently employed in NAP-XPS. Both the identity of the scattering medium as well as two experimental parameters, the gas-phase pressure p and the sample-to-aperture distance d, were chosen at random. The gas-phase pressure p was taken from an interval of [0.1 mbar, 5.0 mbar] and the distance d between the sample and the aperture was taken from [0.1 mm, 2.0 mm]. Finally, we normalized the spectrum in the [0, 1] intensity range. A pseudocode description of the data set generation is shown in Algorithm S1 in the supporting information. For each data set that was used in this work, 250 000 synthetic XP spectra were simulated to create a fully unbiased data set of artificial linear combinations of the chosen reference spectra. For the data shown in this work, artificial single-element data sets were created based on reference spectra of Co, Cu, Fe, Mn, Ni, Pd, and Ti. Additionally, datasets containing multiple elements were created based on reference spectra spanning the whole 2p regions of Ni, Co, and Fe.

Due to the underlying physical processes, Auger excitations for a given chemical species are generally visible at a constant kinetic energy, while core-level spectra exhibit a distinct constant binding energy position. The scale for binding energy E_B is related to the measured kinetic energy E_{kin} and the energy of the incident X-ray photons $h\nu$ according to $E_B = h\nu - E_{kin}$. Therefore, depending on the X-ray excitation energy used during the experiment, Auger spectra may be shifted on the binding energy scale, such that they overlay with the core-level spectra of the same or of different chemical species. This effect typically occurs during synchrotron-based XPS measurements, since the photon energy can be readily tuned by upstream monochromator mirrors. In order to account for this effect, we randomly selected whether or not Auger spectra were additionally used during the linear combination. If Auger spectra were considered, we first randomly selected which chemical species shall contribute to the spectrum and then added both core-level and Auger spectra of these phases to the simulation. The Auger spectra were randomly positioned on the binding energy axis, which accounts for changes in the excitation energy. The scaling parameters of the Auger and core-level spectrum during the linear combination were set to be the same, in order to take into account that both spectra are created by excitation of the same atoms in the material.

In Fig. 1, we show a schematic of the training set simulation workflow, in this case for core-level iron reference spectra (metallic Fe, FeO, Fe_3O_4, and Fe_2O_3, Fig. 1(a)). The four panels in the middle illustrate the data augmentation steps described above (Fig. 1(b–e)). In Fig. 1(f), some examples of mixed metal-oxide iron spectra are shown. It is evident that using the methods described above, it is possible to create iron spectra that exhibit a diverse range of compositions and line shapes, similar to measured spectra from real samples. Therefore, the simulation method described here is suitable for creating data for training neural networks that can later be for inference on 'real' data, with the advantage of explicitly knowing the "ground truth" labels during training and validation.

Along with the spectral data, the atomic concentration of each species was saved as labels for the neural network training. Two different types of labels have been considered. In a most straightforward way, each spectrum could be labeled using the random stoichiometry parameters that were used for the linear combination of reference spectra. However, we considered that the intensity of the Auger and core-level spectrum for each phase stem from the same chemical species and therefore their contribution should only be counted once. So, if both the Auger spectrum and the core-level spectrum of a particular species each contributed 35% of the spectral intensity during simulation, the label for that phase was also just 35% (and not 70%, as in the straightforward labeling described above). Therefore, the final label is an array of N numbers $\bar{y}_i \in [0.1, 1.0],\ i = 1 \ldots N$, which directly represent the relative elemental quantification for all N phases.

Fig. S2 shows the average concentration of each species in a simulated iron mixed metal-oxide data set based on the core-level reference spectra of metallic Fe, FeO, Fe_3O_4, and Fe_2O_3. In each of the data sets (training, validation and test data), the average stoichiometry is close to 25% for each species, which means that across all simulated spectra, the chemical states are distributed at random. Therefore, the simulated data can be considered fully unbiased with respect to the represented stoichiometry.

2.3. Neural network design

The simulated data sets were then used to train convolutional neural networks for automatic quantification. A schematic of the architecture of these convolutional neural network is shown in Fig. 2. The network comprises of a hybrid geometry, with the convolutional layers acting as feature extractors and the fully-connected layers acting as a classification block. The data passed to the input layer of the neural network is the one-dimensional spectral intensity, which comprises of 840–2400 data points, depending on the range and energy step size of the measured reference spectra that were used as bases for the simulation. The data is

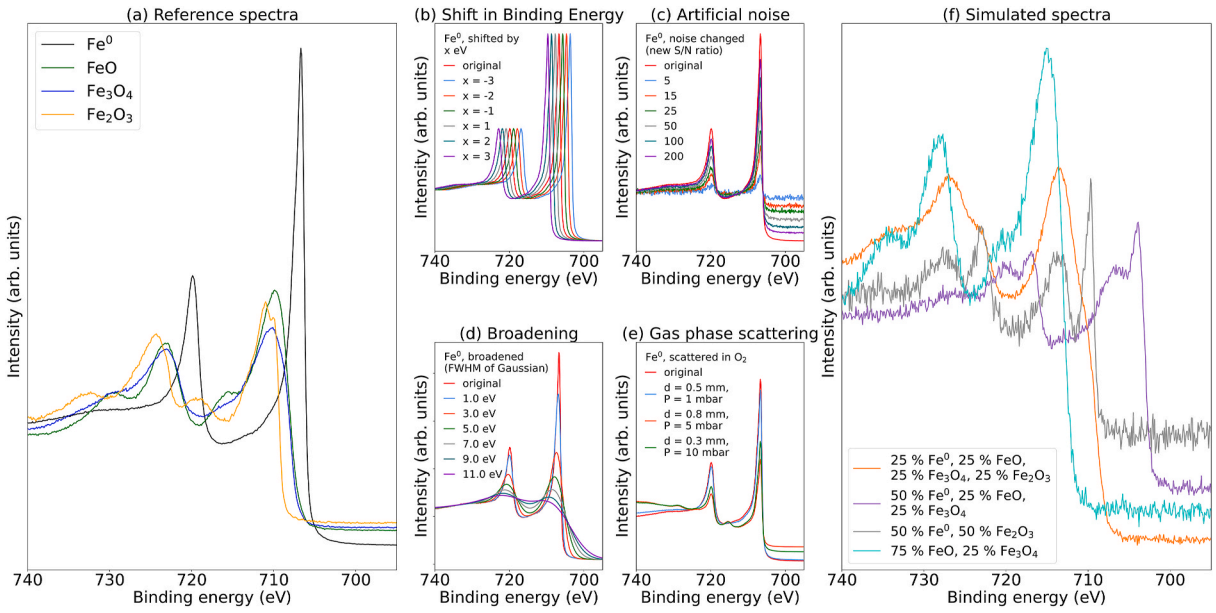


Fig. 1. Schematic of the training set simulation workflow. Well-known reference spectra (in this case, from metallic Fe and three Fe oxides) are linearly combined using random scaling parameters (a). Afterwards, several data augmentation steps are used: (b) shifts along the energy axis, (c) application of artificial Poisson-distributed noise, (d) convolution with a Gaussian to change the energy resolution and (e) scattering in a gas phase (here: O_2). (f) Shows some examples of the simulated mixed metal oxide Fe spectra with different concentrations of each reference.

then passed to the feature extraction part of the network, which comprises of three parallel 1D convolutional layers (each with 12 filters and stride length 1) with varying kernel sizes of $k = 5$, 10, 15, respectively. These parallel convolutional layers with different kernel sizes have been shown to enable the network to learn from small, medium, and long range correlations in the spectra [18].

The three parallel convolutional layers are concatenated and their output is passed to two additional 1D convolutional layers with 12 filters each. The kernel size of these layers is 15 and the stride length is 1. Dropout with a rate of 0.2 is used for regularization for each of the convolutional layers. Afterwards, the resulting data are average pooled (kernel size 2, stride length 2), flattened and passed to the second part of the network, the quantification stage. The quantification stage consists of two fully connected layers. The first fully connected layer has 4000 nodes, while the last layer has N nodes, where N is the number of independent chemical species represented in the data set. All layers except for the last one are using the Rectified Linear Unit (ReLU) as the activation function, while the last layer uses a Sigmoid activation function (logistic function of the form $S(x) = 1/(1 + e^{-x})$), which enables the output of continuous values (regression). All convolutional and fully-connected layers were initialized used the default Glorot uniform initializer. After the second fully-connected layers, the output is normalized in the [0, 1] range. The final output of the neural network is the normalized intensity y_i (with $\sum_i^N y_i = 1$) of all chemical species in the material.

2.4. Model training

In order to properly train the neural network models on the simulated data sets, the correct output activation and loss functions had to be chosen. Here, it is of importance that (a) we want the network to produce continuous outputs for each label (which means that we are dealing with a regression problem), and (b) labels are not independent, but need to sum to 1 since they represent the relative atomic percentage of all phases. We achieved this by using a Sigmoid activation function in the second-to-last layer (yielding continuous results) and a non-trainable final layer, which normalized the outputs such that they sum to one.

We repeatedly trained the same neural network architecture on a relatively simple data set (containing only linear combinations of metallic Ni and NiO reference spectra) and a more challenging data set (based on the relatively similar reference spectra for MnO, Mn_2O_3 and MnO_2), using three different loss function: the mean absolute error (MAE, also called L_1 loss), the mean squared error (MSE, also called L_2 loss), and a custom modified L_2 loss function which scales the L_2 norm by the net results squared. While previous research on CNNs for XPS data suggest that this custom loss function can act as a 'high-pass filter' and therefore can be beneficial for achieving high-accuracy for large concentrations of species [18], the best results for both data sets were achieved using the comparatively 'simple' MAE loss. Therefore, all subsequent models were trained using the MAE loss which is defined as:

$$\mathrm{MAE} = \mathscr{L}(y, \bar{y}) = \frac{\sum_{i=1}^{N} |y_i - \bar{y}_i|}{N}, \tag{2}$$

where y_i is the network output and $\bar{y}_i$ the target values for the i-th label. N is the number of output values of the neural network, which varies from 2 (for Ni and Pd) up to 4 (for Fe and Ti), depending on the number of reference spectra that were used to generate the data set that the neural network model is trained on. Note that in this work, the MAE is calculated not from the atomic percentages, but from the decimal values, meaning that if a spectral label implies 25% of a spectrum arises from a given species, the input to the MAE is 0.25.

Prior to model training on a particular data set, the data set was randomly split into training, validation and test subsets. 70% and 10% of spectra were assigned to the training and validation set, respectively, while the remaining 20% of spectra formed the test set. The models were trained with a batch size of 32 using an ADAM optimizer [36]. The learning rate was kept constant at 1×10^{-5}, while the other parameter of the optimizer were set to their default values ($\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\varepsilon = 1 \times 10^{-8}$). Model training was terminated after 500 epochs since the neural networks did not show any significant performance improvements anymore (stagnation of the training loss) and in some cases, the validation loss became much higher than the training loss after training for more than 700 epochs (overfitting). Even though some models did show improvements in the training loss after more than 500 epochs, these models were already so well optimized after 500 epochs that even the worst predictions differed from the true labels by much less

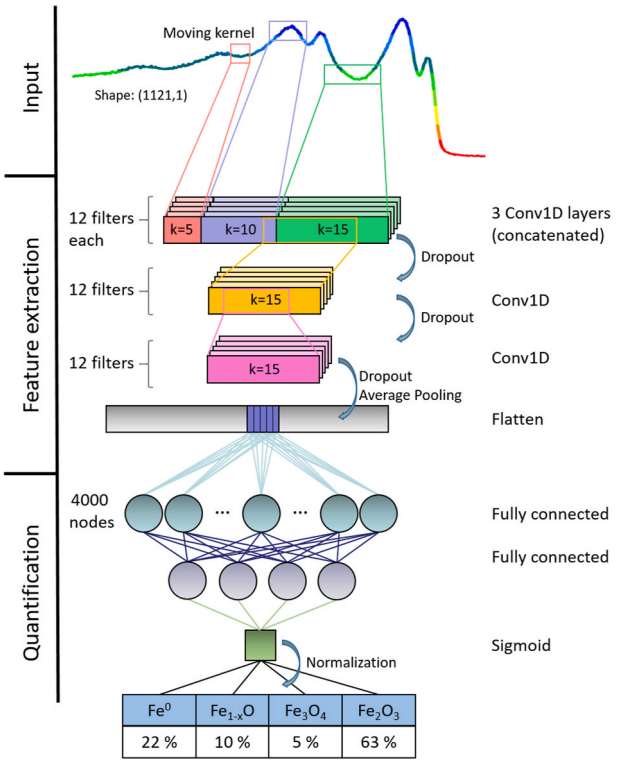


Fig. 2. Schematic of the neural network architecture. In a forward pass, the input data is passed through three one-dimensional convolutional layers of variable kernel size k with 12 filters each ($k = 5, 10, 15$). The resulting kernels are concatenated and passed through two additional one-dimensional convolutional layers (12 filters with a kernel size of 12). During a forward pass, dropout with a rate of 0.2 is applied in each convolutional layer. After this feature extraction module, the data is flattened, average-pooled (pool size = 4) and passed through two fully-connected dense layers. The first dense layer has 4000 nodes, while the second one has N nodes, where N is the number of species that the network is learning to quantify (in this case, $N = 4$). In the last layer, a Sigmoid activation function is used for quantification, i.e. regression. Finally, the output is normalized in the [0, 1] range.

compared to what could reasonably be detected in an XPS measurement considering the finite signal-to-noise ratio. The training time for 500 epochs was around five to 7 h, depending on the number of points in each spectrum in the data set.

2.5. Performance metrics

To evaluate the performance of the trained neural networks as well as other quantification approaches yielding output quantifications $\mathbf{y} = (y_1, \ldots, y_N)$, several metrics are used throughout this work to compare these quantifications to the ground-truth labels $\bar{\mathbf{y}} = (\bar{y}_1, \ldots, \bar{y}_N)$:

- **Mean Absolute Error**: As described above, the mean absolute error describes the average absolute difference between the predictions y_i and labels $\bar{y}_i$ and is defined as

$$\mathrm{MAE} = \frac{\sum_{i=1}^{N} |y_i - \bar{y}_i|}{N}. \tag{3}$$

Aside from being used as the loss function during training, the MAE is also employed as a performance metric in this work, with lower MAE values representing a closer fit of the model to the data. Due to the fact that the sum of all labels is 1 for all spectra, the maximum MAE (which indicates the worst fit possible) for a spectrum is $\frac{2}{N}$, where N is the number of different compounds that make up the label.

- **Maximum Absolute Error**: While the MAE contains information about the *average* absolute difference between the predictions y_i and labels $\bar{y}_i$, we may also be interested in how much the predictions for the *individual* phases differ from the correct stoichiometry. In XPS literature, the error bars for quantification often represent a percentage on the individual species. Therefore, we define the Maximum Absolute Error as

$$\mathrm{MaAE} = \max_i |y_i - \bar{y}_i|. \tag{4}$$

Note that the maximal value of the MaAE is 1 since all labels are normalized in the [0, 1] range.

With these metrics, we can define values for both MAE and MaAE which can act as arbitrary cutoffs for determining a fit as 'correct'. For example, if we declare quantifications correct if the associated MAE is below 0.1, this means that the average prediction on all labels shall not differ by more than 10% from the actual label. If instead we declare quantifications correct if the MaAE is below 0.1, this imposes the stricter condition that each of the predictions for the *individual* phases shall not differ by more than 10%. Given a metric m, we can then calculate the percentage P of a set of quantifications Q_j that are below threshold value t and are therefore considered 'correct':

$$P(Q_j | m, t) = \frac{Q_j(m < t)}{Q_j(m)} \cdot 100\%. \tag{5}$$

These metrics thus allow us to compare the performance of different quantification methods in an intuitive way, comparable to how the error is typically reported in XPS literature.

2.6. Computational hardware and software, data and code availability

The design of the model architectures as well as the training and testing of the neural networks were performed using TensorFlow and Keras in Python [37]. Model training occurred on a GPU (NVIDIA Tesla K80, 24 GB RAM) on Google Colab [38]. The forward and subsequent backward pass through the neural took about 30–50 s for each epoch, viz. for all 160 000 spectra in the training data set. The notebooks used for training and predictions as well as the code for simulation, training, and plotting are available in a GitHub repository [39]. JSON files containing the parameters used during simulation are available in this repository as well. Full data sets as well as the trained TensorFlow models were too large to be uploaded to GitHub, but are available on the KEEPER archive of the Max Planck society [40]. Peak fitting of XP spectra was performed with the CasaXPS software package [41].

3. Results and discussion

3.1. Training on artificial datasets

Fig. 3 shows the evolution of the training and validation loss (mean absolute error) during the training of convolutional neural networks on training sets containing simulated spectra for seven different transition metals and their oxides (Co, Cu, Fe, Mn, Ni, Pd, and Ti). The L_1 loss (MAE) converges towards 0 for all of the elements, with virtually no difference between training and validation. It is therefore safe to assume that the neural network architecture outlined above is suitable since the data are fitted well without any overfitting. Interestingly, the learning of the different neural networks proceeded at varying rates. The neural network trained on the data set containing simulated Mn spectra (based on the reference spectra for MnO, Mn_2O_3, and MnO_2) learned much slower than the CNN trained on simulated Pd spectra (metallic Pd, PdO). This difference in the learning process can be explained by the fact that for each of the elements, the different oxidation states exhibit different amounts of variability in their line shapes and peak positions. Neural networks that were trained on elements for which reference spectra of different phases have higher similarity (like for the Mn references) took

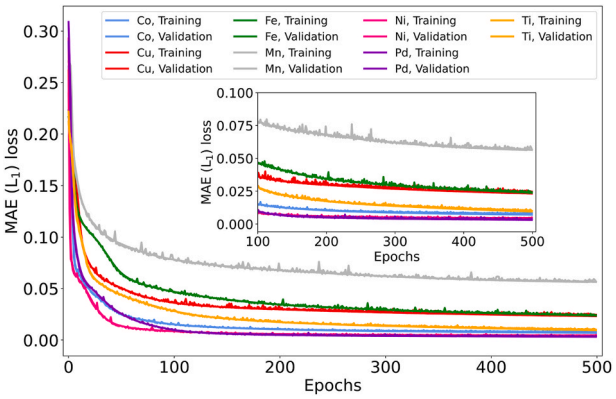


Fig. 3. Evolution of the training and validation loss (mean absolute error) for multiple neural networks as a function of the number of forward passes (epochs) through artificial data sets containing the main core-level regions of the transition metals indicated in the legend. The inset shows a zoomed-in version of the same plot, focusing on the loss after the 100th epoch.

considerable longer times to converge than those trained on elements with less similar reference states (like Pd), as will be discussed in the next paragraph.

It is evident from the decrease in training and validation loss that the neural network is learning to correctly predict the quantification for a given spectrum. However, in order to see how well the models generalize to new data and when and why they fail to do so, the models need to be tested on out-of-sample test data. Fig. 4 shows some examples for neural network quantifications on spectra from the test data sets. In Fig. 4(a–c), examples are shown (in green) for which the quantifications (shown as inset tables) almost perfectly agree with the labels from the simulation. Fig. 4(a) shows an example from the iron test data set, while examples for Mn and Ti are displayed in Fig. 4(b) and (c), respectively.

For all of these examples, the MAE loss is below 0.01 and the maximum difference between prediction and ground truth for any individual species is below 1%, well below what is typically achievable by traditional peak fitting methods. In Fig. S3 in the supporting information, histograms of the mean absolute error between the simulated stoichiometry and the neural networks' outputs are shown for the complete training sets for each of the transitional metals. For the neural network models that have a higher average loss after 500 epochs of training (such as the models for Cu and Mn), the distribution of the MAEs has a longer tail towards higher values, indicating that the neural network's ability to correct quantify some of these spectra is limited. For the elements with fewer input reference spectra that are more easily distinguishable (such as Ni or Pd, not shown here), the median MAE is much lower and the histogram does not show such a prolonged tail.

In order to examine why the neural network models fail for some spectra in the test data sets, the spectra with the highest associated loss were investigated. In Fig. 4(d–f), some examples are shown (in red) for which the quantification deviates substantially from the simulated stoichiometry. From these examples, we can identify a few causes for the breakdown of predictive strength of the neural networks. Fig. 4(d) shows an example where the simulated gas-phase scattering so strongly influences the spectrum that the line shape is radically different compared to the reference spectra, with a strong increase of intensity towards the high binding energy side. While it is technically possible to encounter such a spectrum (which simulates scattering in N_2 gas at a pressure of 4.3 mbar and an aperture-to-sample distance of 2.0 mm), it seems unreasonable that a trained spectroscopist would be able to achieve a better fitting result than the one that the neural network trained on iron data produced. Therefore, we consider this an edge case for which it is expected that the neural network does not produce useable quantifications. Fig. 4(e) shows another example for quantification on a mixed iron spectrum. In this case, the signal-to-noise ratio (which was simulated to be 2.5) is so low that it is essentially impossible to distinguish the different iron species. Again, when using typical peak

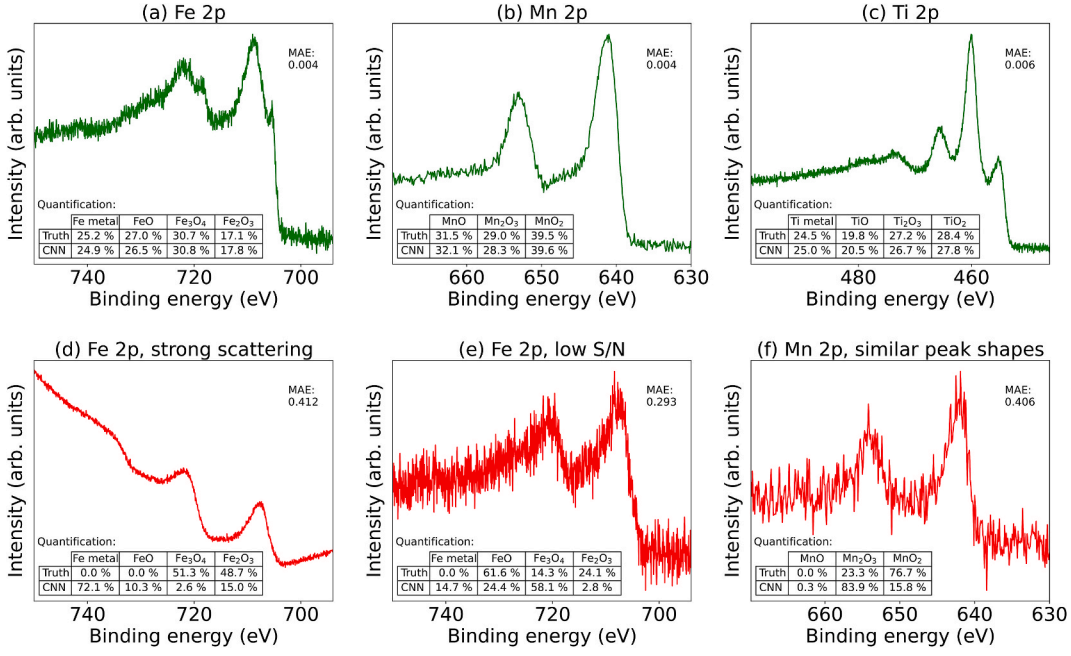


Fig. 4. Examples for quantifications of CNNs on artificial test datasets containing mixed Fe (a, d, e), Mn (b, f) and Ti (c) spectra. The spectra in green (a–c) show results where the quantification obtained from the CNN can be considered to be correct. As a cutoff for this figure, quantifications were deemed 'good' if the associated MAE is in the lowest 10% of MAEs obtained on the test data set and if the individual atomic percentages for each species do not differ by more than 4% compared to the values used for the linear combination during data set simulation. The spectra in red (d–f) show examples where the neural network approach fails, i.e., where the MAE is in the highest 10% of MAEs obtained on the test data set and where the atomic percentages for each species differ by more than 4% compared to the values used for the linear combination during data set simulation. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

fitting approaches, it is unreasonable to expect better resemblance between the simulated and predicted labels. Therefore, this example effectively demonstrates the strength of this neural network approach: it shows high overlap of prediction and ground truth for spectra which are relatively easy to quantify, while the network's ability declines for spectra with higher embedded ambiguity (in this case, the presence of high statistical noise). A third example of a non-successful prediction of a trained neural network is shown in Fig. 4(f), for a spectrum based on reference spectra for different manganese oxides. While the signal-to-noise ratio of this spectrum (5.8) is higher than for the iron example, the neural network fails at distinguishing between Mn_2O_3 and MnO_2. The reason for this failure can be found when examining the reference spectra of these two Mn oxide (Fig. S1(d)). The Mn 2p spectra of pure Mn_2O_3 and MnO_2 are already rather similar, such that the differentiation between them is challenging for the neural network and for a human analyst alike. Therefore, it is expected that the prediction fails even for a higher signal-to-noise ratio compared to the other elements. This effect also explains why the neural network trained on Mn data has the highest average loss after 500 epochs and why the MAE histogram shows a longer tail towards higher values. For the histogram in Fig. S3(d), all MAEs above 0.15 arise for spectra that have strong contributions of Mn_2O_3 and MnO_2 and are either significantly broadened or exhibit a high noise compared to the signal. The example of Mn therefore shows the limit of predictability that this convolutional neural network can achieve. The same effect can explain why the Cu CNN has a longer tail towards higher MAEs since line shapes of metallic Cu and Cu_2O are very similar in BE position and shape. It should be noted however that the MAE for 81.6% of the simulated Mn spectra is below 0.1 (which translates to an average deviation of prediction and ground truth of $\sim$ 10%), so that for the majority of Mn spectra, the neural network produces reasonable results. Indeed, an example of a good prediction on a simulated mixed Mn oxide spectrum is shown in Fig. 4(b). For spectra with this level of noise (S/N: 21.0) and Gaussian broadening (FWHM: 1.06 eV), the neural network achieves a deviation of less than 1% for each species (MaAE $<$ 0.01).

In order to investigate the effect that the size of the training set has on the training performance, additional training runs for the models were performed using subsets of different sizes (25 000, 50 000, 100 000, 150 000, 200 000, and 250 000 spectra) of the both the simulated Ni and Mn data sets. In Fig. S4, the effect of different subset sizes on both the evolution of the training and validation losses as well as the test loss after 1000 epochs is shown for neural networks trained on simulated Ni 2p and Mn 2p spectra. It is evident that increasing the training data size leads to a better performance both during training (i.e., the training loss decreases faster) as well as during inference after 1000 epochs, with a lower test loss for neural networks that were trained on more data. No saturation effect was detected for bigger training sets.

3.2. *Effects of the simulation parameters on the neural network performance*

In the previous section, the effect of noise and Gaussian broadening was discussed qualitatively. In order to obtain quantitative information about how the spectral parameters affect the neural networks' ability to predict the correct quantifications, we performed a more in-depth analysis for the neural network that was trained on simulated mixed Mn oxide spectra. Fig. 5 shows plots of three simulation parameters against the MAE obtained by comparing the neural network outputs with the simulated stoichiometries.

Due to the inherent translational invariance of convolutional neural networks [42], it is expected that the shift along the binding energy axis does not affect the output of the neural network. Indeed, the points in Fig. 5(a) do not show a trend in any direction, suggesting that the shift of the simulated spectra along the binding energy axis does not change the performance of the neural network. The Pearson correlation coefficient of shift and MAE is just -0.01. Note here that the absolute values of the shift are discretely distributed, since they are multiples of the energy step sizes of the input reference spectra.

For the signal-to-noise ratio (Fig. 5(b)), a trend emerges such that lower S/N corresponds to a worse performance of the neural network (higher MAE). S/N ratio and MAE are anti-correlated (Pearson correlation coefficient: -0.40). For a S/N ratio above 10, most of the data is still fit reasonably well, with 80% of MAEs below 0.1. However, for very low S/N ratios, the network performs much worse. Since a trained spectroscopist would also have much lower quantification accuracy on noisy data, it is expected that the neural network would show a similar behavior. However, even for very low S/N ($<$ 5), the quantification obtained by the neural network is still reasonably good for 48% of the

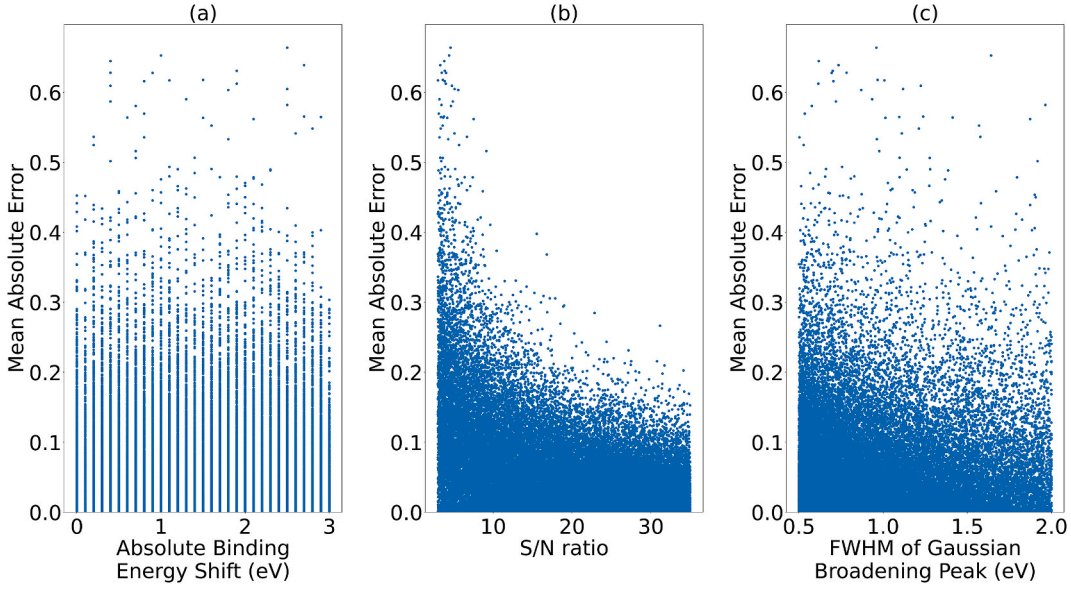


Fig. 5. MAE losses obtained by using the Mn neural network on simulated spectra from the Mn test set as a function of different simulation parameters: (a) the absolute shift across the binding energy axis, (b) the signal-to-noise (S/N) ratio with respect to the simulated Poisson noise, and (c) the Full Width at Half Maximum (FWHM) of the Gaussian peak used for simulating spectral broadening. In all three plots, each point represents the MAE of the NN prediction on one of the 40 000 spectra in the Mn test data set.

spectra, with a MAE below 0.1.

The third parameter under investigation is the FWHM of the Gaussian broadening peak, which was allowed to vary between 0.5 eV and 3.0 eV during the simulation (Fig. 5(c)). For the broadening, there is no clear trend as the MAE of the network output does not change with the FWHM of the Gaussian peak. The Pearson correlation coefficient of 0.01 is very close to 0, so the FWHM does not influence the MAE of the network's prediction very much. Indeed, the median MAEs of the neural network predictions for very broadened spectra (FWHM $>$ 2.5 eV) and spectra that were not broadened significantly (FWHM $<$ 1.0 eV) are identical (0.025). It seems that the neural network is quite robust against Gaussian broadening effects that can occur due to the XPS instrument and its setting. As a further test, we also allowed the FWHM to vary up to 10 eV (not shown here). When allowing for very high FWHMs, the neural network's performance was stable up to a FWHM values of $\sim$ 7.5 eV. For even stronger broadening, the spectra were dominated by the Gaussian peak and did not show many distinguishable features anymore, which led to the degradation of the network's performance and higher mean absolute errors during inference.

3.3. Validation against traditional peak fitting methods

The previous sections showed that the neural network approach can work for spectra that were not part of the training data set. While it would be desirable to show that the network models can also be used to quantify "real" spectra, i.e., spectra that were measured rather than simulated, this runs into a logical dilemma. In order to show that the neural network can accurately quantify real spectra, we would need to know the 'ground truth' of these spectra, meaning that the correct quantification should be known beforehand. However, as outlined above, even through careful peak fitting, the results depend strongly on the methods used, such that the quantification errors often exceed 10%. Since the simulated test spectra represent a fully unbiased data set of artificial linear combinations of the chosen reference spectra, it is reasonable to assume that any quantification method that works well for these spectra would also work well for measured spectra (assuming that these consist of linear combinations of the same references as well). Therefore, in order to determine the accuracy of the neural network models, we compared its performance on simulated spectra from the test data set to three different methods of peak fitting that are typically used in the literature.

An exemplary simulated spectrum from the iron test data set is shown in Fig. 6(a), along with its 'ground truth' quantification, viz. the simulated stoichiometry. One of the most cited approaches for fitting transition metal XP spectra is the one by Biesinger, Grosvenor et al. [43,44]. In this approach (called method M1 here), only the main core-level XPS peak and its background (e.g. for Fe 2p, the Fe $2p_{3/2}$ region) are fitted with peaks constrained to conform to multiplets obtained from Hartree-Fock calculations. The fit for the Fe 2p spectrum from the test data set, based on the parameters given in Grosvenor et al. [43], is shown in Fig. 6(b). In this case, the possible species were limited to metallic Fe, FeO, Fe_3O_4, and Fe_2O_3. For comparison, we have also adopted the often-used approach of building a curve-fitting model for each of the reference spectra by manually tuning peak parameters (line shapes, positions, widths) until the residual between the fitting envelope and the reference spectrum is minimized. Afterwards, these tightly-constrained peak models are combined to fit a mixed metal-oxide spectrum (Fig. 6(c), method M2). Finally, another commonly used technique is to extract the line shapes of the reference spectrum of each pure chemical state after background removal, and then using these line shapes for fitting the spectrum based on the same background function (method M3) [1,45]. This approach is shown for the same Fe spectrum in Fig. 6(d), using a U 2 Tougaard background [46]. A more detailed description of each peak fitting approach is available in the supporting information material. For each of the three approaches, the obtained quantification as well as the mean absolute error with respect to the ground truth is shown as inset in the corresponding figure.

Each of the different methods of fitting the Fe 2p spectrum returns a slightly different quantification for the four iron species, as indicated in the insets in Fig. 6. The neural network on the other hand returns the following quantification: Metallic Fe: 29.9%, FeO: 16.9%, Fe_3O_4: 23.4%, and Fe_2O_3: 29.8%, which is much closer to the actual quantification from the simulation (see Fig. 6(a)). In order to determine the validness of each of the quantification approaches outlined above, a random subset

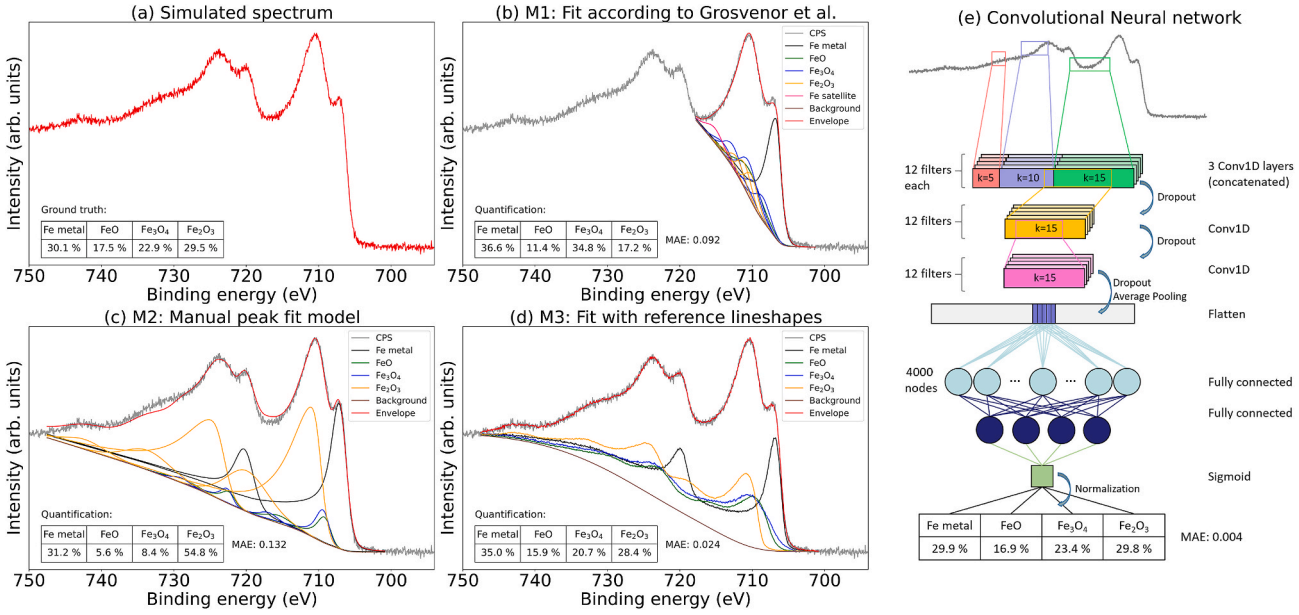


Fig. 6. Validation of neural network model against traditional peak fitting methods. (a) Simulated Fe 2p spectrum based on reference spectra for metallic Fe, FeO, Fe_3O_4, and Fe_2O_3. The parameters that were used for the linear combination are shown in the inset table. (b) Peak fitting of the Fe $2p_{3/2}$ region, based on Grosvenor et al. [43] (c) Peak fitting based on a peak model obtained for each reference spectrum. (d) Peak fitting using the line shapes extracted from reference spectra of the Fe compounds (based on a U2 Tougaard background.) (e) Scheme of the neural network trained on simulated Fe 2p spectra. The tables in (b–e) show the quantification obtained through the different methods.

of 100 simulated Fe 2p spectra from the test data set was fitted. Fig. S5 shows histograms of mean absolute errors (MAE) achieved for the three manual peak fitting methods (a-c), along with a histograms of MAEs achieved by the convolutional neural network (d). It is immediately obvious that the neural network outperforms each of the conventional peak fitting method with the lowest median MAE of 0.015. The lower and upper quartiles are found at 0.005 (−0.010 vs. median) and 0.032 (+0.017 vs. median), respectively. Interestingly, the median MAE is an order of magnitude smaller than the variation of both the spectral intensities (mean standard deviation of 0.10 relative to the average intensity for each energy over all test spectra) and the labels (mean standard deviation of 0.28 relative to the average label value for species over all test spectra). For the fitting approaches, the median MAEs are much higher than for the neural network: 0.131 (lower quartile: 0.010, upper quartile: 0.178) for method M1 [43,44], 0.184 (lower quartile: 0.111, upper quartile: 0.257) for method M2 (manual peak fitting), and 0.090 (lower quartile: 0.043, upper quartile: 0.0154) for method M3 (based on reference line shapes from pure phases), respectively.

While it is interesting to see that the neural network's output produced the lowest mean absolute error and thus the quantification with the highest degree of correctness, the average error in itself does not tell us how many of the quantifications can be seen as correct in a practical sense. In order to determine how many of the spectra were classified correctly, we chose an arbitrary, yet practical, cutoff level below which we consider the quantification correct. Taken into account the experimental uncertainty as well as the uncertainty that is expected from any valid method for XPS quantification [1,5], we chose to consider quantifications valid where the average point difference between the determined atomic percentages of the species was within 10% of the stoichiometry used during simulation. This cutoff is equivalent to a MAE of 0.1 (see Equation (3)). With this cutoff, the neural network was able to correctly classify 99 of the 100 spectra (99%). All of the fitting approaches performed worse, since under this cutoff value, only 26% (method M1), 21% (M2), and 54% (M3) of spectra were correctly quantified, respectively. If one restricts the cutoff such that for *each* of the species the difference between ground truth and predictions is below 10% (which corresponds to a MaAE of 0.1, see Equation (4)), the number of correctly quantified spectra dropped a bit for the neural network (92%), but even further for the three peak fitting approaches (11% (method M1 based on Ref. 43, 44), 9% (M2, manual peak fitting), and 38% (M3, based on extracting pure phase line shapes)). It is noteworthy that the line shape-based approach together with a Tougaard background produced the most accurate results of all the fitting approaches, but still was outperformed by the neural network.

3.4. Quantifying multiple spectral regions at the same time

Up to this point, we have only considered materials where the whole sample consisted of metal and oxide species from a singular element. However, in real XP spectra, there is commonly more than one element present. In order to represent this fact, we have additionally trained CNNs on spectra containing multiple elements. As an example, we considered materials containing Ni, Co, and Fe. For these three elements, the quantification cannot be performed on each of the core-level 2p regions by itself since there is a significant overlap of Auger peaks with the core-level 2p spectra when using a conventional Al Kα X-ray source (with a photon energy of $h\nu = 1486.61$ eV). For example, all Co species have a strong LMM Auger signal at a kinetic energy of 765–775 eV, which for the Al Kα source overlaps strongly with the Fe 2p region. In order to show how strongly the Auger peaks from the other materials are disturbing the core-level 2p spectra of each element, we measured each species in the range of 880 eV–695 eV, which includes the 2p region of each species, but also the main LMM Auger lines which overlap with the core-level peaks of the other elements. The reference spectra are shown in Fig. S6 in the supporting information. It is obvious from these spectra that in order to correctly quantify spectra with multiple elements, we must take the contributions of Auger transitions into account.

Therefore, we designed new training data sets that contain a linear combination of all Ni, Co, and Fe species across the whole Ni2p - Co2p - Fe2p spectral regions (namely, metallic Ni, NiO, metallic Co, CoO, Co_3O_4, metallic Fe, FeO, Fe_3O_4, and Fe_3O_4). This approach allows for the Auger peaks to be naturally incorporated into the data set. The input passed to the neural network model then consisted of spectra in the energy range of the whole Ni 2p - Co 2p - Fe 2p spectral region, while the output labels are the nine chemical species listed above. For this model, the training time was extended to 1300 epochs, as it took longer for the neural network to converge. Fig. 7 shows the training and validation loss of this new model (a) as well as some exemplary spectra from the withheld test set (b-e). From these plots, it is apparent that the CNN is able to correctly quantify the presence of all nine chemical species even with the presence of Auger peaks.

In order to compare the results achieved from the neural network to more traditional peak fitting methods, we performed a fit based on line shapes extracted from the same phase, in a similar manner as in method M3 described above (Fig. 8(a)). The quantification obtained from the peak fitting and the neural network are shown as an inset table in Fig. 8, along with the true labels, i.e., the stoichiometry used during simulation. It is clear that the neural network outperforms the peak fitting method. Note that for these complex spectra, even the line shapes-based method, which produced the best results for single-element spectra, exhibits significant differences between the actual line shape and the fitting envelope. We also tried to fit such spectra with method M2 described above, but the residual of the fit with respect to the measured data was even bigger.

Again, in order to compare the neural network to the peak fitting model, a random subset of 100 simulated spectra from the test data set containing the nine reference spectra in the Ni 2p - Co 2p - Fe 2p spectral regions was fitted using method M3. The histograms of MAE losses of the peak fitting and neural network approaches are shown in Fig. 8(b). For this data set, the neural network outperformed the peak fitting method once again, as the median MAE of 0.027 (for the CNN) was ~ 1.5 times lower than the median MAE of 0.041 for the line shapes-based method. The histograms in Fig. 8(b) show that this performance advantage can mostly be traced to the fact that the neural network was able to perfectly quantify the stoichiometry for a substantial subset of the spectra. It is notable that histograms of both the neural network and the fitting approach show a long tail towards higher losses (around MAE = 0.04). These higher losses occur mostly on spectra that contain multiple different iron species, especially iron oxides. This effect is also visible in the two examples in Fig. 7(d and e), which show spectra that were not correctly quantified by the neural network. The fact that the presence of multiple iron oxide species leads to worse performance results can be explained by two reasons: for one, the main Fe 2p regions are very similar for the different Fe oxides, while the Co and Ni reference spectra are easier to distinguish in their respective 2p regions. Moreover, the LMM Auger spectra of FeO, Fe_3O_4, and Fe_2O_3 are quite similar (see Fig. S6), so that it is not possible to distinguish between the iron oxides based on the shape of the Auger spectra, which are located in the Co 2p and Ni 2p regions when using an Al Kα source.

For the purpose of having a practical descriptor of whether the quantifications can be considered correct, we choose the same cutoff as for the single-element spectrum, namely that 'valid' quantifications are those for which the average difference between the simulated and determined stoichiometry across all chemical states is below 10% (MAE $\leq$0.1). With this threshold, the neural network's quantification can be considered correct for all 100 spectra. With the peak fitting approach, only 80% of the spectra are correctly quantified. If we limit ourselves to the stricter cutoff that each *individual* stoichiometry must not differ by more than 10% (MaAE $\leq$0.1), the neural network still reaches 71% accuracy, while the peak fitting approach can only correctly quantify 64% of spectra.

With peak fitting approaches, it is difficult to produce valid

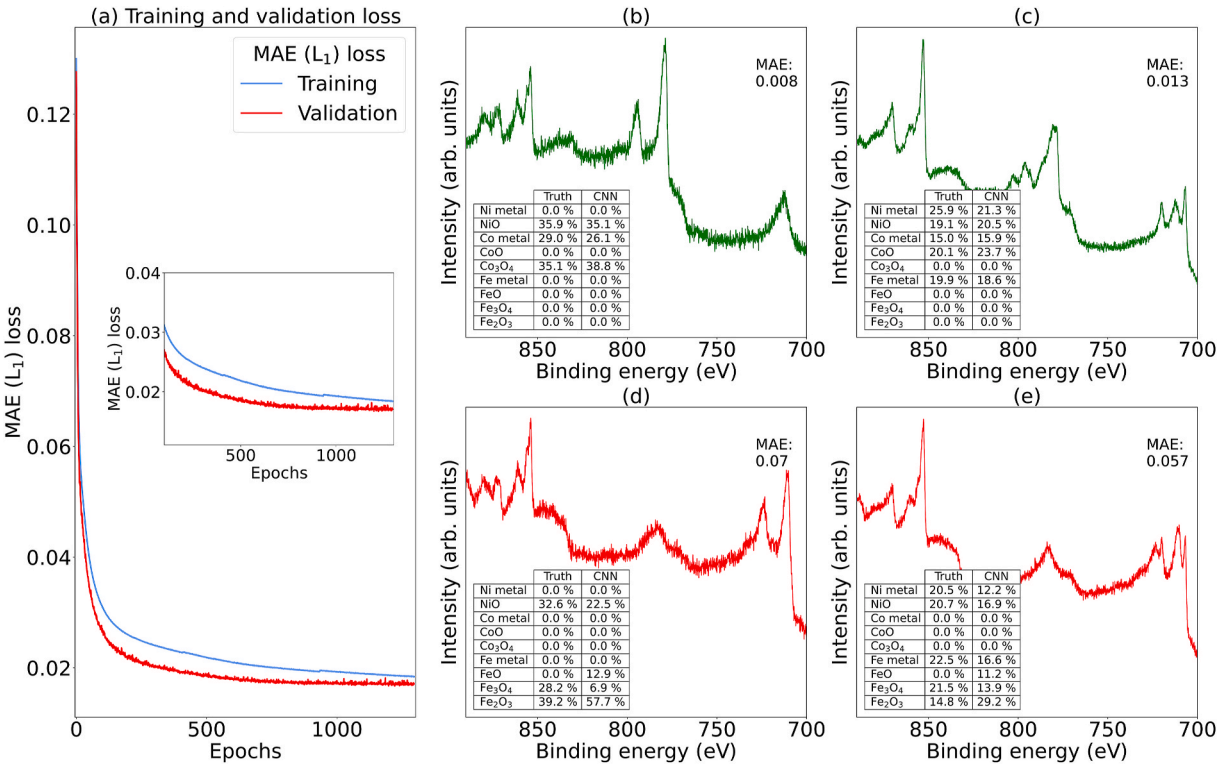


Fig. 7. (a) Evolution of the training and validation loss (mean absolute error) for a neural network during training on a mixed Ni/Co/Fe data set as a function of the number of forward passes (epochs). The inset shows a zoomed-in version of the same plot, focusing on the loss after the 100th epoch. (b–e) Examples for quantifications of CNNs on artificial test datasets containing mixed Ni/Co/Fe spectra. The spectra in green (b,c) show results where the quantification obtained from the CNN can be considered correct. As a cutoff for this figure, quantifications were deemed "good" if the associated MAE is in the lowest 10% of MAEs obtained on the test data set and if the individual atomic percentages for each species do not differ by more than 4% compared to the values used for the linear combination during data set simulation. The spectra in red (d,e) show examples where the neural network approach fails, that is, where the MAE is in the highest 10% of MAEs obtained on the test data set and where the atomic percentages for each species differ by more than 4% compared to the values used for the linear combination during data set simulation. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

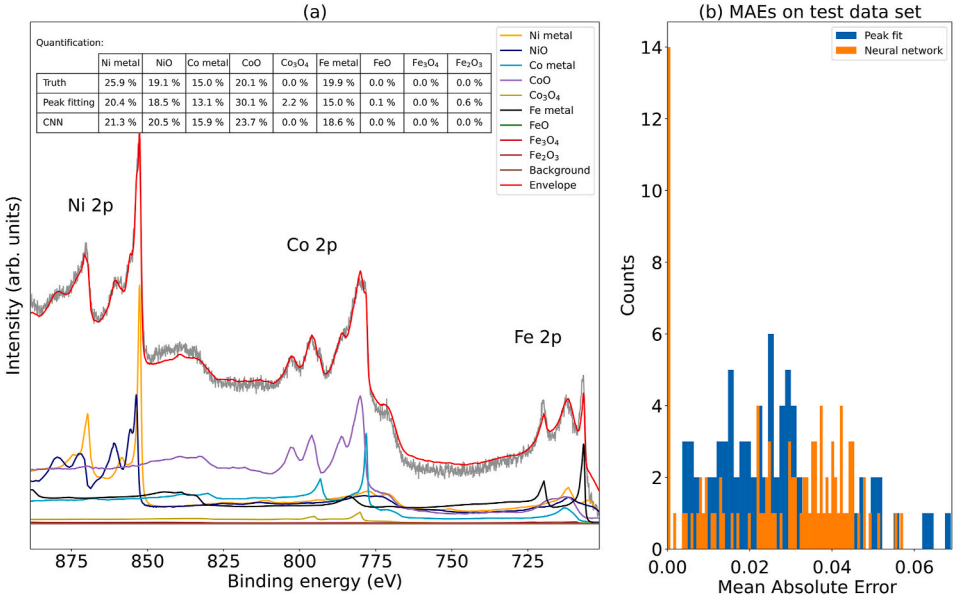


Fig. 8. (a) Simulated spectrum of the Ni 2p, Co 2p and Fe 2p regions of a mixed Ni/Co/Fe oxide sample. Peak fitting was performed by using the line shapes of the indicated species across the whole binding energy range (890 eV–684 eV). The inset table shows a comparison of the quantifications that were obtained from peak fitting and from a CNN that was trained on simulated mixed Ni/Co/Fe spectra with the same binding energy range, along with the 'ground truth' quantification, viz. the simulated stoichiometry. (b) Histogram of MAE losses of the peak fitting (orange) and CNN (blue) approaches for all spectra in the simulated Ni/Co/Fe test data set. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

quantifications for complex spectra containing multiple elements with overlapping Auger peaks. The simulation and neural network framework presents an opportunity to quantify the presence of different chemicals states with much higher accuracy.

3.5. *Allowing for variable input and step sizes*

While the results in the previous sections show that it is possible to use the CNN approach to quantify spectra of single XPS regions as well as multiple concatenated regions, the neural networks used so far exhibit

the limitation that the input sizes are fixed, which means that the binding energy range as well as the step size need to be the same for any new spectrum which is to be passed to the model. This can be problematic since the binding energy range is typically not the same across multiple experiments. Therefore, it is desirable that the models should work for spectra of different lengths. One option may be to artificially shorten or extend the spectrum such that it has the same range as the artificial set that was used for training. However, extending the spectrum, which may be achieved by continuing the last data points on both BE sides, can significantly perturb the results of the neural network. Shortening the spectrum may results in loss of features that are critical for quantification.

To circumvent the issue of having different binding energy ranges, we took a different approach. While the datasets remained the same as before, instead of training on complete spectra, we randomly selected sub-regions of each spectrum (with a defined width) and trained similar models as before on these subsets. The approach is schematically shown in Fig. 9(a). In this example, for each spectrum that is passed to the neural network model, only a random region of 100 eV width is used. Therefore, the input layer of the model has $I = w \cdot s$ nodes, where w is the size of the binding energy window and s is the step size.

Fig. 9(b) shows the training loss of a model that was trained on a data set that contained a linear combination of all Fe and Co species across the whole Co 2p - Fe 2p spectral regions. In this case a random window of $w = 100$ eV was used to select a smaller part of the spectrum that was passed to the model during training. In Fig. 9(c), the quantification on the spectrum from the withheld training set is shown. Even though the window that was used during quantification is far away from the Fe 2p region, the model trained on the random windows of the training set is still able to accurately quantify the different Fe species due to the presence of Fe Auger lines in the window region (i.e., in the Co 2p region).

We then used this neural network with $w = 100$ eV to perform inference on the test set containing spectra from the combined Fe 2p and Co 2p regions after choosing random windows of 100 eV of each spectrum. Fig. 9(d) shows a histogram of the MAE between the CNN's output and the simulated stoichiometries. The median MAE was 0.050, while the lower and upper quartiles are found at 0.030 (−0.020 vs. median) and 0.067 (+0.017 vs. median), respectively. While these MAE values are higher compared to the model that was trained on the complete spectra (see previous section), the MAE on 95% of the spectra was below the previous established cutoff of 10%, while the quantifications on 52% of the spectra still satisfied the criterion of MaAE < 0.1 (meaning that for each species, the difference between ground truth and CNN prediction was below 10%). With a cutoff value of MaAE < 0.15, 64% of spectra were still correctly quantified. Therefore, this "window" approach still produces valid quantifications for a majority of spectra.

With the "window" approach, a single model trained on data sets that contain the species that one is interested in can be used for spectra with varying binding energy ranges and step sizes. This approach can be especially useful when one wants to obtain a quantification for large datasets of XP spectra (e.g., for a spectrum database or a high-throughput experiment) with various different spectral parameters, without having to manually change each spectrum such that it fits to the classifier model.

3.6. Quantifying prediction uncertainty

In the previous sections, we demonstrated that CNNs can be used to accurately quantify chemical species even in complex XP spectra in an efficient manner. However, it may also be desirable to not just predict the atomic concentration of each species, but also how uncertain the network is about its prediction. From a spectroscopy perspective, this may, for instance, be useful when additional phases are present in the

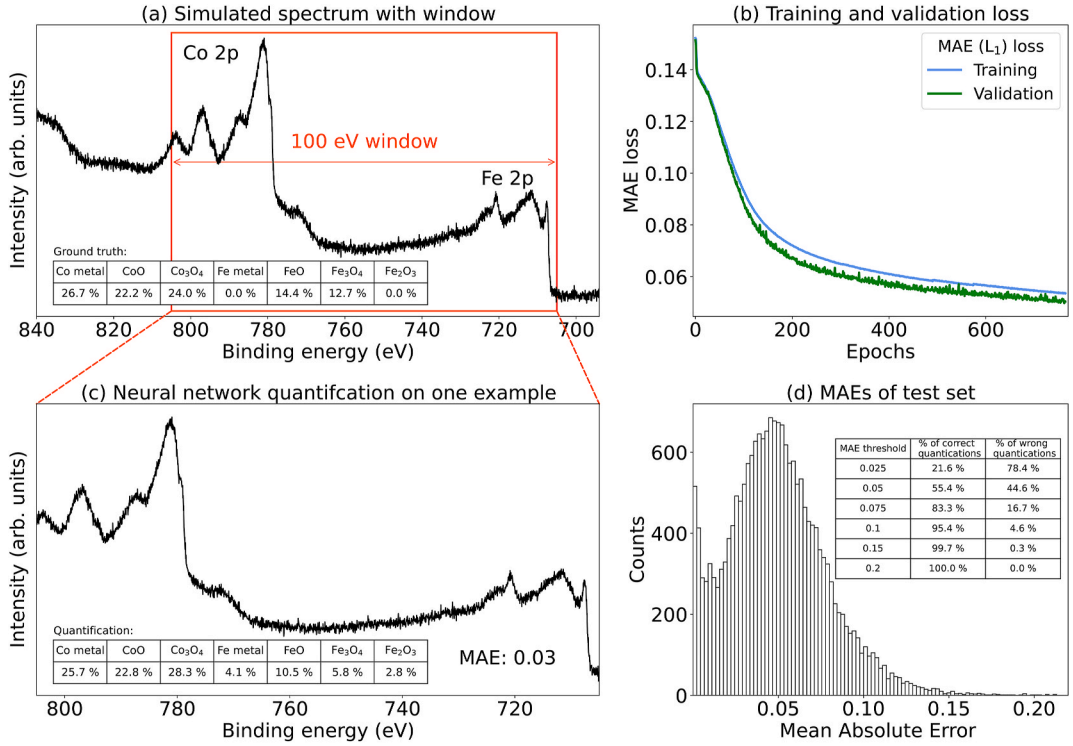


Fig. 9. (a) Selection of random 'sub-spectrum' of the XP spectrum that is then passed to the model during training and inference. (b) Training loss of a model that was trained on 100 eV 'sub-regions' of spectra from a data set that contained a linear combination of all Fe and Co species across the whole Fe 2p - Co 2p spectral region. (c) Example for the quantification of a spectrum from the withheld training set. The inset table shows the quantification using the neural network model trained on spectra with a binding energy range of 100 eV. (d) Histogram of MAEs for all spectra in the test data set. The inset table shows how many of the CNN quantifications can be considered correct, assuming a given MAE as a threshold for correctness.

spectrum that were not in the training spectra. As an example, if an Fe 2p spectrum contains iron sulfide or nitride phases, the network outlined in the previous sections (which only quantified metallic and oxide species of iron) should return a prediction as usual (attempting to extrapolate from the training data), but should additionally return some quantity conveying a high level of uncertainty with such an input spectrum, signaling that it cannot correctly classify the spectrum. If implemented in such a way, the uncertainty could then be used as an indicator that the spectrum in question may contain additional or unusual species, which may warrant higher attention and a more detailed analysis by an XPS expert. This could be particularly useful when the network is used to quantify spectra in large data sets (e.g., in databases or during high-throughput measurements), since it may be unfeasible to perform a detailed analysis of every spectrum in the data set. Another possible desired behavior could be that strong distortions in the spectrum (e.g., due to strong scattering or due to a very low signal-to-noise ratio, see Fig. 4) would lead to a decrease in the quantification confidence and prediction of a high uncertainty.

Uncertainty evaluation in neural networks is an active research field known under the theme of *Bayesian Deep Learning*. The most principled approach is to train probabilistic models from data using Bayesian inference. In a Bayesian neural network, all model parameters (weights and biases) are represented by probability distributions. During training, the probability distribution of each parameter is learned in a way that coherently explains the variability in the training data. More rigorously, the idea is to place a prior distribution $p(\delta)$ (Bayesian prior) on the weights and biases and learn a posterior distribution $p(\delta|D)$ (Bayesian posterior) over the parameters δ of a probabilistic model given some observed data D using Bayes theorem as:

$$p(\delta|D) = \frac{p(D|\delta)p(\delta)}{p(D)} = \frac{p(D|\delta)p(\delta)}{\int p(D|\delta)p(\delta)d\delta}. \quad (6)$$

During inference, multiple forward passes through the network are performed, each time with a new set of parameters sampled from $p(\delta|D)$. The resulting distribution of predictions yields a representation of both the variability in the training data and the model uncertainty. While these Bayesian models offer a mathematically grounded framework to reason about model uncertainty, they typically require rigorous hyperparameter tuning and take longer to converge. Indeed, when we replaced the parameters in our CNN models with probability distributions and tried to retrain them using the well-known *Bayes by Backprop* approach [47], the model did not converge at all, possibly because the choice of Bayesian prior and training hyperparameters (both of which are not trivial) was not appropriate.

An alternative to the mathematically rigorous treatment outlined above is the so-called Dropout Variational Inference method. Under the non-Bayesian approach to deep learning, dropout layers are typically used for regularization during network training [48]. Dropout works by randomly setting a pre-specified percentage of network parameters to 0 during training, thus allowing the network to learn a more robust representation of the data and preventing overfitting. For the Bayesian approach, the idea is to not just use dropout during training, but also at test time, i.e., during the forward pass of a test spectrum through the network. Hence, every single pass through the network will result in a different prediction. It can be shown that the stochastic distribution of the predictions obtained by multiple forwards passes approximate the posterior distribution $p(\delta|D)$ in Bayesian models [49]. The implementation of our uncertainty-aware neural network can thus be reduced to performing dropout not only during training, but also during inference. The neural networks that we trained on the transition metal XP spectra already contained a dropout layer. Therefore, if we simply use these trained models and, at test time, we evaluate the stochastic model output, we can approximate the predictive posterior and thus the uncertainty of the network's output. This process is also called Monte Carlo (MC) dropout.

In Fig. 10, we show this approach for a neural network model trained on Ni 2p spectra (with outputs of metallic Ni and NiO). Fig. 10(a) shows two simulated Ni 2p spectra (blue, orange). Spectrum S1 (in blue) represents an example of a well-defined spectrum, with a high signal-to-noise ratio (S/N = 16.3) and very little broadening (FWHM of Gaussian of 1.0 eV). Spectrum S2 (orange), on the other hand, has significantly higher noise (S/N = 1.4) and broadening (FWHM = 2.4 eV). Along with these simulated spectra that are similar to those in the training data set, another spectrum is shown (green, spectrum S3), which was created by adding two Gaussian peaks to the first spectrum (at BE = 857.5 eV and 876.5 eV, respectively, both with a FWHM of 0.75 eV). These additional peaks represent a species the neural network has never seen during training and can therefore be used to estimate how uncertain the network's prediction are on data which is out of the scope of its original purpose.

In Fig. 10(b) and (c), histograms of 1000 predictions on each spectrum are shown. Fig. 10(b) shows the histogram for the prediction of the metallic Ni content, while the prediction of the NiO content is shown in 10(c). For the well-defined spectrum S1, which should be easy to quantify, the network's prediction across 1000 forward passes with activated dropout does not significantly vary. On average, the neural network predicted atomic concentrations of (67.5 ± 1.0) % for metallic Ni and (32.5 ± 1.0) % for NiO, respectively (for the linear combination during simulation, the concentrations were 67.6% for metallic Ni and 32.4% for NiO, respectively). Thus, the prediction exhibits a high confidence (or, equivalently, a low uncertainty). For the noisy spectrum S2, the histogram of predictions is much broader and the average prediction of (62.8 ± 4.9) % for metallic Ni and (32.6 ± 4.9) % for NiO, respectively, has a bigger standard distribution. While the neural network does predict the 'ground truth' values of 66.5% (metallic Ni) and 33.5% (NiO) on average (within the prediction error), it is less certain about its prediction, which is reflected in the broader distribution of the predicted

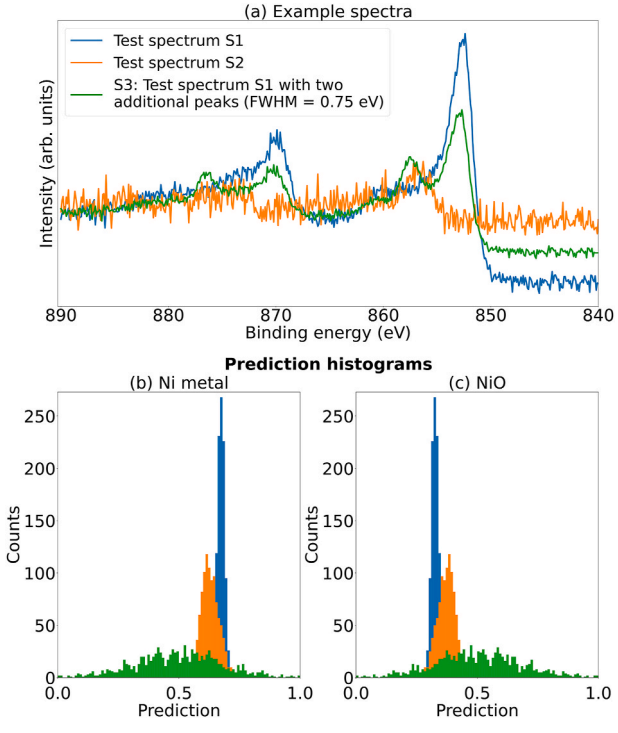


Fig. 10. (a) Two simulated, realistic Ni 2p spectra (S1, S2) from linear combinations of reference spectra for metallic Ni and NiO, along with a synthetic 'toy' spectrum (S3). (b–c) Histogram of quantification predictions of the neural network for metallic Ni (a) and NiO (b). During each forward pass of the spectra, dropout was implemented to obtain a Monte Carlo sampling of different predictions.

quantifications. For noisy spectra with more broadened features, the network exhibits a lower confidence.

For the spectrum that was out of the scope of the initial training set (S3), it is observed that the prediction certainty is much lower as well. For spectrum S3, the prediction of the neural network is essentially randomly distributed across each forward pass, since it is impossible for the network to extract any features from the spectrum. The most abundant prediction is 50% for both species, which is just the learned average across the whole training data set. This shows that if a new spectrum contains any phases that the network did not encounter during training (represented by the additional generic Gauss peaks here), the model does predict the known phases, but with a very high uncertainty. This may alert the user of the network to the fact that this spectrum was unusual and may need to be fitted with additional species.

Using Dropout Variational Inference, we were able to show that it is possible to obtain a measurement of the uncertainty associated with the neural network's quantification. Most notably, the network's prediction becomes more uncertain if there are any unknown species in the spectrum. This may be very useful when large datasets of XP spectra are quantified. If the prediction on a particular spectrum is associated with a high uncertainty, this spectrum may warrant more detailed analysis by an expert spectroscopist.

3.7. Approximations and limitations

While the previous sections showed clearly that the CNN framework can be a valuable tool for the automatic quantification of transition metal XP spectra, it is important to understand the assumptions and approximations that are made in the simulation as well as in the network, so that one can understand the limits of applicability of this approach:

1. The most obvious limitation is that during the simulation the spectra of reference materials are linearly combined. This explicitly assumes that there is a homogeneous mixing of the different phases on the surface of the material under investigation. Therefore, if a neural network model that was trained on these simulated data sets is then applied for inference on new spectra, it will only ever produce the correct quantification if the different phases on the surface are mixed homogeneously as well. Specifically, no depth distribution is taken into account which is always critical in XPS due to its inherent surface sensitivity. In principal, more complex combinations of the reference spectra could be simulated as well (rather than a simple linear combination), with the concentration of different phases depending on their localization in the sample. In this case, the neural network may even be able to output the spatial morphology of the material. However, while certainly feasible, a data set for training such a model would have to contain a large amount of different surface configurations and would therefore require a very different simulation process, which was not part of the scope of this work.
2. The output options of the neural network are entirely dependent on the reference spectra that are used during the simulation of the training data set. Naturally, it follows that any chemical species that was not part of the data set cannot be quantified. For example, a neural network trained on data containing metallic iron and the three iron oxide phases outlined above, will never be able to predict the presence of different iron compounds (like iron nitride or sulfide). If the user suspects that additional phases may be present, the network would have to be retrained on a new data set that also contains these phases. Note that, typically, it may not be necessary to retrain the entire network. It may be sufficient to fix the parameters of the feature extraction stage and only retrain the fully-connected layers in the quantification stage. This approach is called transfer learning [50,51]. However, this limitation of the phases that can be predicted can also be used to the user's advantage. As we showed in the last section, the inherent Bayesian character of the trained networks (due to the dropout layers) allows predictions about the quantification uncertainty and thus, indirectly, about whether the phases that the networks were trained on are present in the test spectrum. Specifically, when an unknown phase is added, the CNNs' predictions are essentially random (Fig. 10). This property may be especially useful in *in situ/operando* NAP-XPS experiments where intermediates in a reaction may only exist under operating conditions and hence there are no suitable reference spectra available. For example, this could be particularly helpful in heterogeneous catalysis research, where the active phase is often a metastable phase which is only available at high pressures and/or temperatures on the particular catalysts and thus the measurement of a reference spectrum is impossible [52]. A high uncertainty prediction of the neural network can then alert the user to the fact that the surface phases in the catalyst are *not* represented by the standard compounds and the spectrum warrants further attention.
3. An expert in XPS analysis would typically not just analyze the features of the main region of the photoemission spectrum, but also consider additional spectral regions such as the O 1s region. Since one objective of the neural network models shown here is to be used in order to quantify large data sets in databases or during high-throughput measurements, we focused on the main region since this is what is typically always available, while many auxiliary measurements of different regions are often not taken. Nevertheless, this is not an inherent limitation, since for more advanced analysis, the networks could be scaled to include multiple regions (as for the combined Fe–Co spectra in section 3.4) or even new CNNs could be trained on reference data for the auxiliary regions. In any case, even XPS experts can profit from the automation the proposed framework provides in order to optimize their workflow. For example, traditionally, the analysis has to be performed for each individual spectrum, which does not scale well to large data sets. Additionally, the quantification obtained through the neural network prediction can act as an independent comparison for the analysis the XPS user performs manually, thus strengthening the confidence in the obtained analysis results.
4. The quantification of multiple spectral regions at the same time as shown in section 3.4 requires the high-resolution measurement of reference spectra for all materials that the user is interested in, across the whole binding energy range of the main XP peaks from every material. Here, we only show this for selected metal and oxide references of Ni, Co, and Fe. If one aims to scale this approach towards even more complex combinations of materials, the measurement time for the reference spectra increases significantly. Therefore, this approach may only work if the XPS user has an underlying assumption of the possible species (e.g., from synthesis or from other characterization techniques) and could therefore limit the amount of full-scale reference spectra to measure.

4. Conclusions

In summary, this work has introduced a framework for using convolutional neural networks for the automatic analysis of transition metal XP spectra. We have demonstrated that our CNN is able to quantify chemical species in such spectra in less than a second, with an accuracy that rivals the traditional approaches for XPS data analysis. Moreover, this approach is flexible enough to scale to spectra containing multiple elements or to datasets of spectra that were measured in different energy ranges and/or different step sizes. One can imagine that within this framework even more complex spectra, resulting, for example, from additional species or from over-layers, could be simulated and a CNN could be trained to quantify these spectra as well. Since the only input to the data set simulation are well-characterized reference spectra and the simulated spectral distortions are commonly seen also in other techniques, we imagine that the simulation and CNN training framework could easily map to other spectroscopic tools where the data is measured

in the energy or frequency domain.

As described in the previous section, the framework communicated here suffers from some approximations and limitations. Therefore, one may suggest that a better quantification could be obtained by a trained spectroscopist who can take all these limitations into account. However, we would like to stress that the objective of this work is explicitly not to supersede the analysis by XPS experts or sophisticated theoretical calculations. Instead, our CNNs may be used by non-expert analysts or for exploratory data analysis in large data sets. Moreover, the CNNs are applicable for situations where many different spectra need to be accurately quantified at once, for example, to enrich the metadata of a database or during high-throughput measurements. However, even the trained spectroscopist can profit from the automation provided through the proposed framework. For example, the results of the typical analysis could be compared to the neural network quantification, thus allowing for a 'tandem' approach of manual and automated CNN-based analysis.

Additionally, we demonstrated that the same neural network models can also be used to not only correctly quantify the chemical species that contributed to the spectrum, but also how uncertain the network is about this quantification. This feature can be particularly helpful to identify spectra where more species than the 'standard' reference compounds are present.

Funding

This work was financially supported by the Max Planck Society for the Advancement of Science, Germany.

Financial disclosure

None reported.

CRediT authorship contribution statement

Lukas Pielsticker: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing - original draft. **Rachel L. Nicholls:** Software, Data curation, Writing – review & editing. **Serena DeBeer:** Resources, Writing – review & editing, Supervision. **Mark Greiner:** Conceptualization, Methodology, Software, Writing – review & editing, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Datasets as well as trained CNN models are available in the KEEPER archive of the Max Planck Society. The code and parameters of the data set simulation are available on GitHub.

Acknowledgments

The authors would like to thank Gudrun Klihm (MPI CEC) for her help in measuring reference XP spectra.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.aca.2023.341433.

References

[1] G.H. Major, N. Fairley, P.M. Sherwood, M.R. Linford, J. Terry, V. Fernandez, K. Artyushkova, Practical guide for curve fitting in X-ray photoelectron spectroscopy, J. Vac. Sci. Technol. A: Vac. Surf. Films 38 (6) (2020) 061203, https://doi.org/10.1116/6.0000377.

[2] P.M. Sherwood, The use and misuse of curve fitting in the analysis of core X-ray photoelectron spectroscopic data, Surf. Interface Anal. 51 (6) (2019) 589–610, https://doi.org/10.1002/sia.6629.

[3] G.H. Major, T.G. Avval, D.I. Patel, D. Shah, T. Roychowdhury, A.J. Barlow, P. J. Pigram, M. Greiner, V. Fernandez, A. Herrera-Gomez, A discussion of approaches for fitting asymmetric signals in X-ray photoelectron spectroscopy (XPS), noting the importance of Voigt-like peak shapes, Surf. Interface Anal. 53 (8) (2021) 689–707, https://doi.org/10.1002/sia.6958.

[4] G. Greczynski, L. Hultman, X-ray photoelectron spectroscopy: towards reliable binding energy referencing, Prog. Mater. Sci. 107 (2020) 100591, https://doi.org/10.1016/j.pmatsci.2019.100591.

[5] G.H. Major, T.G. Avval, B. Moeini, G. Pinto, D. Shah, V. Jain, V. Carver, W. Skinner, T.R. Gengenbach, C.D. Easton, Assessment of the frequency and nature of erroneous X-ray photoelectron spectroscopy analyses in the scientific literature, J. Vac. Sci. Technol. A: Vac. Surf. Films 38 (6) (2020), 061204, https://doi.org/10.1116/6.0000685.

[6] C.R. Brundle, B.V. Crist, X-ray photoelectron spectroscopy: a perspective on quantitation accuracy for composition analysis of homogeneous material, J. Vac. Sci. Technol. A 38 (4) (2020), 041001, https://doi.org/10.1116/1.5143897.

[7] A.G. Shard, Practical guides for X-ray photoelectron spectroscopy: quantitative XPS, J. Vac. Sci. Technol. A: Vac. Surf. Films 38 (4) (2020), 041201, https://doi.org/10.1116/1.5141395.

[8] M.H. Engelhard, D.R. Baer, A. Herrera-Gomez, P.M. Sherwood, Introductory guide to backgrounds in XPS spectra and their impact on determining peak intensities, J. Vac. Sci. Technol. A: Vac. Surf. Films 38 (6) (2020) 063203, https://doi.org/10.1116/6.0000359.

[9] N.A. Belsey, D.J.H. Cant, C. Minelli, J.R. Araujo, B. Bock, P. Brüner, D.G. Castner, G. Ceccone, J.D.P. Counsell, P.M. Dietrich, M.H. Engelhard, S. Fearn, C. E. Galhardo, H. Kalbe, J.W. Kim, L. Lartundo-Rojas, H.S. Luftman, T.S. Nunney, J. Pseiner, E.F. Smith, V. Spampinato, J.M. Sturm, A.G. Thomas, J.P.W. Treacy, L. Veith, M. Wagstaffe, H. Wang, M. Wang, Y.-C. Wang, W. Werner, L. Yang, A. G. Shard, Versailles project on advanced materials and standards interlaboratory study on measuring the thickness and chemistry of nanoparticle coatings using XPS and LEIS, J. Phys. Chem. C 120 (42) (2016) 24070–24079, https://doi.org/10.1021/acs.jpcc.6b06713.

[10] H. Shinotsuka, H. Yoshikawa, R. Murakami, K. Nakamura, H. Tanaka, K. Yoshihara, Automated information compression of XPS spectrum using information criteria, J. Electron. Spectrosc. Relat. Phenom. 239 (2020) 146903, https://doi.org/10.1016/j.elspec.2019.146903.

[11] M. Suzuki, H. Nagao, Y. Harada, H. Shinotsuka, K. Watanabe, A. Sasaki, A. Matsuda, K. Kimoto, H. Yoshikawa, Raw-to-repository characterization data conversion for repeatable, replicable, and reproducible measurements, J. Vac. Sci. Technol. A 38 (2) (2020) 023204, https://doi.org/10.1116/1.5128408.

[12] R. Murakami, H. Tanaka, H. Shinotsuka, K. Nagata, H. Shouno, H. Yoshikawa, Development of multiple core-level XPS spectra decomposition method based on the Bayesian information criterion, J. Electron. Spectrosc. Relat. Phenom. 245 (2020) 147003, https://doi.org/10.1016/j.elspec.2020.147003.

[13] R. Murakami, H. Yoshikawa, K. Nagata, H. Shinotsuka, H. Tanaka, T. Iizuka, H. Shouno, Automatic estimation of unknown chemical components in a mixed material by XPS analysis using a genetic algorithm, Sci. Technol. Adv. Mater.: Methods 2 (1) (2022) 91–105, https://doi.org/10.1080/27660400.2022.2061878.

[14] S.-H. Park, H. Park, H. Lee, H.-S. Kim, Iterative peak-fitting of frequency-domain data via deep convolution neural networks, J. Kor. Phys. Soc. 79 (12) (2021) 1199–1208, https://doi.org/10.1007/s40042-021-00346-1.

[15] J. Acquarelli, T. van Laarhoven, J. Gerretzen, T.N. Tran, L.M. Buydens, E. Marchiori, Convolutional neural networks for vibrational spectroscopic data analysis, Anal. Chim. Acta 954 (2017) 22–31, https://doi.org/10.1016/j.aca.2016.12.010.

[16] D. Chen, Z. Wang, D. Guo, V. Orekhov, X. Qu, Review and prospect: deep learning in nuclear magnetic resonance spectroscopy, Chem. Eur J. 26 (46) (2020) 10391–10401, https://doi.org/10.1002/chem.202000246.

[17] M. Chatzidakis, G.A. Botton, Towards calibration-invariant spectroscopy using deep learning, Sci. Rep. 9 (1) (2019) 2126, https://doi.org/10.1038/s41598-019-38482-1.

[18] G. Drera, C.M. Kropf, L. Sangaletti, Deep neural network for X-ray photoelectron spectroscopy data analysis, Mach. Learn.: Sci. Technol. 1 (1) (2020), 015008, https://doi.org/10.1088/2632-2153/ab5da6.

[19] M. Gallagher, P. Deacon, Neural networks and the classification of mineralogical samples using X-ray spectra, in: Proceedings of the 9th International Conference on Neural Information Processing, 2002. ICONIP'02, vol. 5, IEEE, 2002, pp. 2683–2687, https://doi.org/10.1109/ICONIP.2002.1201983.

[20] K. Ghosh, A. Stuke, M. Todorović, P.B. Jørgensen, M.N. Schmidt, A. Vehtari, P. Rinke, Deep learning spectroscopy: neural networks for molecular excitation spectra, Adv. Sci. 6 (9) (2019) 1801367, https://doi.org/10.1002/advs.201801367.

[21] J. Liu, M. Osadchy, L. Ashton, M. Foster, C.J. Solomon, S.J. Gibson, Deep convolutional neural networks for Raman spectrum recognition: a unified solution, Analyst 142 (21) (2017) 4067–4074, https://doi.org/10.1039/C7AN01371J.

[22] C.M. Pate, J.L. Hart, M.L. Taheri, RapidEELS: machine learning for denoising and classification in rapid acquisition electron energy loss spectroscopy, Sci. Rep. 11 (1) (2021) 19515, https://doi.org/10.1038/s41598-021-97668-8.
[23] C.D. Rankine, M.M.M. Madkhali, T.J. Penfold, A deep neural network for the rapid prediction of X-ray absorption spectra, J. Phys. Chem. 124 (21) (2020) 4263–4270, https://doi.org/10.1021/acs.jpca.0c03723.
[24] J. Timoshenko, D. Lu, Y. Lin, A.I. Frenkel, Supervised machine-learning-based determination of three-dimensional structure of metallic nanoparticles, J. Phys. Chem. Lett. 8 (20) (2017) 5091–5098, https://doi.org/10.1021/acs.jpclett.7b02364.
[25] J. Timoshenko, H.S. Jeon, I. Sinev, F. Haase, A. Herzog, B. Roldan Cuenya, Linking the evolution of catalytic properties and structural changes in copper-zinc nanocatalysts using operando EXAFS and neural-networks, Chem. Sci. 11 (2020) 3727–3736, https://doi.org/10.1039/D0SC00382D.
[26] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, G. Carleo, Neural-network quantum state tomography, Nat. Phys. 14 (5) (2018) 447–450, https://doi.org/10.1038/s41567-018-0048-5.
[27] C. Zheng, K. Mathew, C. Chen, Y. Chen, H. Tang, A. Dozier, J.J. Kas, F.D. Vila, J. J. Rehr, L.F.J. Piper, K.A. Persson, S.P. Ong, Automated generation and ensemble-learned matching of X-ray absorption spectra, npj Comput. Mater. 4 (1) (2018) 12, https://doi.org/10.1038/s41524-018-0067-x.
[28] W. Smekal, W. Werner, C. Powell, Simulation of electron spectra for surface analysis (SESSA): a novel software tool for quantitative auger-electron spectroscopy and X-ray photoelectron spectroscopy, Surf. Interface Anal. 37 (11) (2005) 1059–1067, https://doi.org/10.1002/sia.2097.
[29] W.S.M. Werner, Simulation of electron spectra for surface analysis using the partial-intensity approach (PIA), Surf. Interface Anal. 37 (11) (2005) 846–860, https://doi.org/10.1002/sia.2103.
[30] A.G. Shard, Detection limits in XPS for more than 6000 binary systems using Al and Mg Kα X-rays, Surf. Interface Anal. 46 (3) (2014) 175–185, https://doi.org/10.1002/sia.5406.
[31] T.G. Avval, E.F. Smith, N. Fairley, M.R. Linford, Why the signal-to-noise (S/N) ratio in X-ray photoelectron spectroscopy (XPS) generally decreases as binding energy increases, August 2019, Vac. Technol. Coat. Mag. Digit. Mag. (2019) 33–35, https://digital.vtcmag.com/12727/19239/index.html.
[32] M. Salmeron, R. Schlögl, Ambient pressure photoelectron spectroscopy: a new tool for surface science and nanotechnology, Surf. Sci. Rep. 63 (4) (2008) 169–199, https://doi.org/10.1016/j.surfrep.2008.01.001.
[33] D. Frank Ogletree, H. Bluhm, E.D. Hebenstreit, M. Salmeron, Photoelectron spectroscopy under ambient pressure and temperature conditions, Nucl. Instrum. Methods Phys. Res. Sect. A Accel. Spectrom. Detect. Assoc. Equip. 601 (1–2) (2009) 151–160, https://doi.org/10.1016/j.nima.2008.12.155.
[34] L. Pielsticker, R. Nicholls, S. Beeg, C. Hartwig, G. Klihm, R. Schlögl, S. Tougaard, M. Greiner, Inelastic electron scattering by the gas phase in near ambient pressure XPS measurements, Surf. Interface Anal. 53 (7) (2021) 605–617, https://doi.org/10.1002/sia.6947.
[35] S. Tougaard, M. Greiner, Method to correct ambient pressure XPS for the distortion caused by the gas, Appl. Surf. Sci. 530 (2020), 147243, https://doi.org/10.1016/j.apsusc.2020.147243.
[36] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprints, https://arxiv.org/abs/1412.6980, 2014.
[37] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G.S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mane, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viegas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, X. Zheng, TensorFlow: large-scale machine learning on heterogeneous distributed systems, arXiv preprints, https://arxiv.org/abs/1603.04467, 2016.
[38] E. Bisong, Google colaboratory, in: E. Bisong (Ed.), Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners, Apress, Berkeley, CA, 2019, pp. 59–64, https://doi.org/10.1007/978-1-4842-4470-8_7.
[39] L. Pielsticker, R. Nicholls, M. Greiner, S. DeBeer, xpsdeeplearning: public release, GitHub repository. https://github.com/lukaspie/xpsdeeplearning, 2023.
[40] L. Pielsticker, R. Nicholls, M. Greiner, S. DeBeer, DeepXPS: Data Sets and Trained Models, KEEPER archive, 2023. https://keeper.mpdl.mpg.de/d/25ebee640ba546228864a.
[41] N. Fairley, CasaXPS Software Ltd, Online, accessed: 2023-01-02, http://www.casaxps.com, 2022.
[42] Y. LeCun, Learning invariant feature hierarchies, in: A. Fusiello, V. Murino, R. Cucchiara (Eds.), Computer Vision – ECCV 2012. Workshops and Demonstrations, Springer Berlin, Heidelberg, Germany, 2012, pp. 496–505, https://doi.org/10.1007/978-3-642-33863-2_51.
[43] A. Grosvenor, B. Kobe, M. Biesinger, N. McIntyre, Investigation of multiplet splitting of Fe 2p XPS spectra and bonding in iron compounds, Surf. Interface Anal. 36 (12) (2004) 1564–1574, https://doi.org/10.1002/sia.1984.
[44] M.C. Biesinger, B.P. Payne, A.P. Grosvenor, L.W. Lau, A.R. Gerson, R.S.C. Smart, Resolving surface chemical states in XPS analysis of first row transition metals, oxides and hydroxides: Cr, Mn, Fe, Co and Ni, Appl. Surf. Sci. 257 (7) (2011) 2717–2730, https://doi.org/10.1016/j.apsusc.2010.10.051.
[45] H. Liu, G. Wei, Z. Xu, P. Liu, Y. Li, Quantitative analysis of Fe and Co in Co-substituted magnetite using XPS: the application of non-linear least squares fitting (NLLSF), Appl. Surf. Sci. 389 (2016) 438–446, https://doi.org/10.1016/j.apsusc.2016.07.146.
[46] S. Tougaard, Practical guide to the use of backgrounds in quantitative XPS, J. Vac. Sci. Technol. A: Vac. Surf. Films 39 (1) (2021), https://doi.org/10.1116/6.0000661.
[47] C. Blundell, J. Cornebise, K. Kavukcuoglu, D. Wierstra, Weight uncertainty in neural network, in: F. Bach, D. Blei (Eds.), Proceedings of the 32nd International Conference on Machine Learning, vol. 37, of Proceedings of Machine Learning Research, PMLR, Proceedings of Machine Learning Research, Lille, France, 2015, pp. 1613–1622, in: https://proceedings.mlr.press/v37/blundell15.html.
[48] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. 15 (56) (2014) 1929–1958. http://jmlr.org/papers/v15/srivastava14a.html.
[49] Y. Gal, Z. Ghahramani, Dropout as a Bayesian approximation: representing model uncertainty in deep learning, in: M.F. Balcan, K.Q. Weinberger (Eds.), Proceedings of the 33rd International Conference on Machine Learning, vol. 48, of Proceedings of Machine Learning Research, PMLR, New York, NY, USA, 2016, pp. 1050–1059, in: https://proceedings.mlr.press/v48/gal16.html.
[50] K. Weiss, T.M. Khoshgoftaar, D. Wang, A survey of transfer learning, J. Big Data 3 (1) (2016) 9, https://doi.org/10.1186/s40537-016-0043-6.
[51] H.-C. Shin, H.R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, R. M. Summers, Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning, IEEE Trans. Med. Imag. 35 (5) (2016) 1285–1298, https://doi.org/10.1109/TMI.2016.2528162.
[52] Y. Han, H. Zhang, Y. Yu, Z. Liu, In situ characterization of catalysis and electrocatalysis using APXPS, ACS Catal. 11 (3) (2021), https://doi.org/10.1021/acscatal.0c04251.