scientific reports

OPEN

Electron energy loss spectroscopy database synthesis and automation of core-loss edge recognition by deep-learning neural networks

Lingli Kong, Zhengran Ji & Huolin L. Xin ✉

The ionization edges encoded in the electron energy loss spectroscopy (EELS) spectra enable advanced material analysis including composition analyses and elemental quantifications. The development of the parallel EELS instrument and fast, sensitive detectors have greatly improved the acquisition speed of EELS spectra. However, the traditional way of core-loss edge recognition is experience based and human labor dependent, which limits the processing speed. So far, the low signal–noise ratio and the low jump ratio of the core-loss edges on the raw EELS spectra have been challenging for the automation of edge recognition. In this work, a convolutional-bidirectional long short-term memory neural network (CNN-BiLSTM) is proposed to automate the detection and elemental identification of core-loss edges from raw spectra. An EELS spectral database is synthesized by using our forward model to assist in the training and validation of the neural network. To make the synthesized spectra resemble the real spectra, we collected a large library of experimentally acquired EELS core edges. In synthesize the training library, the edges are modeled by fitting the multi-Gaussian model to the real edges from experiments, and the noise and instrumental imperfectness are simulated and added. The well-trained CNN-BiLSTM network is tested against both the simulated spectra and real spectra collected from experiments. The high accuracy of the network, 94.9%, proves that, without complicated preprocessing of the raw spectra, the proposed CNN-BiLSTM network achieves the automation of core-loss edge recognition for EELS spectra with high accuracy.

Ionization edges (core-loss edges) on the electron energy loss spectra (EELS) carry important information not only reflecting the elemental information but also allowing quantitative composition analysis. In addition, the edge onset shift and the near-edge fine structure can be used to deduce the bonding state[1]. Compared to Energy Dispersive X-ray Spectroscopy (EDS), EELS can deliver electronic state information of a material and has higher sensitivity in detecting light elements[2,3]. However, the core-loss region of EELS is also featured by the low signal–noise ratio (SNR). Thus, precisely recognizing the elemental edges is the first step before any advanced analysis.

The traditional way of edge recognition involves multiple steps. Preprocessing operations, including background removal and sometimes multiple-scattering effect deconvolution, are conducted to extract the edges from the spectrum. Afterward, these edges are classified by referring to the database generated by verified experimental data[4]. Therefore, this process requires significant manual labor and is experience based.

Automatically finding ionization peaks in Energy Dispersive X-ray Spectroscopy (EDS) is now routine in commercial software because the peaks are distinctively above the background and the shapes are almost universally gaussian. However, automatically identifying edges in the EELS spectrum has been traditionally difficult because the edges are sitting on a large background, and in most cases, the jump ratio is low. Due to this nature, it has been non-trivial to locating EELS edges automatically. A number of approaches have been proposed to extract edges of unknown components in the spectrum, such as the spatial variant difference approach, multivariate statistical analysis[5] (independent components analysis (ICA)[6,7], principal components analysis (PCA)[8]), etc. Unfortunately, these approaches have constraints and are rarely used in existing EELS software because they are not always robust. Artificial intelligence (AI) has drawn the attention of researchers to simplify complex workflows and reduce labor work. Machine learning algorithms have been successfully implemented in mineral recognition and classification from Raman spectra[9]. A dual Autoencoder-Classifier algorithm was also utilized to achieve the core-loss region denoising[10]. For detecting core-loss edges from the raw EELS spectra, Chatzidakis,

Department of Physics, University of California, Irvine, CA 92617, USA. ✉email: huolin.xin@uci.edu
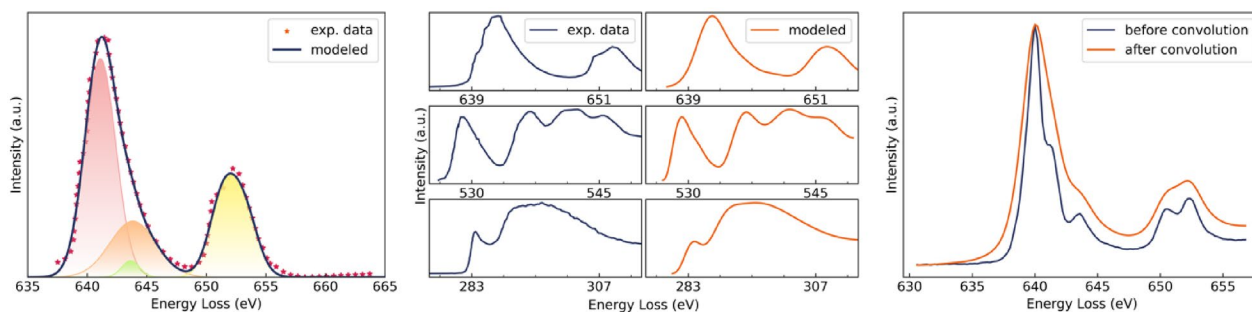
(a) multi-gaussian fitting of Mn-L edge (b) original and modeled edges   (c) Mn-L edge with convolution

Figure 1. Ionization edge modeling.

et. al. proposed a convolutional neural network (CNN) for feature extraction and compound classification, but it was specifically trained for Mn element valance classification[11]. A network that can recognize all the elements on a raw spectrum remains to be developed.

Given that the ionization energy of a specific element is highly related to its electron configuration, nuclear charge, etc., its core-loss edge can only appear within a certain range along the energy loss axis. In other words, the core-loss edges of different elements appear almost "sequentially" on a spectrum. In addition, core-loss edges are present on the plasmon tail and the tails of lower energy core edges[12,13]. Thus, the edge shape is also energy loss dependent. As a result, the traditional convolutional neural network (CNN), which is originally designed for image classification, fails to classify the different core-loss edges on a raw spectrum. The reason being that a pure CNN ended with a fully connected classification layer that is only capable of classifying whether a certain feature is present in the signal/image; it, however, does not use the spatial information, although the information is encoded in the convolutional layers. For example, the latter transition metals, Mn, Fe, Co, Ni all have similar "rabbit-ear" shaped L2,3 edges. A pure CNN network would have the tendency to misclassify these edges because their locations on the energy axis are not used. That is why the region-based CNNs with localization/bounding box prediction layers have been developed to perform both classification and object localization.

In this respect, the recurrent neural networks (RNNs) utilize the correlation in the "time" domain and have been proven to be capable of handling sequence prediction problems. Based on this consideration, we built a Convolutional-Bidirectional Long Short-Term Memory (CNN-BiLSTM) network by combining a CNN network with an RNN network, i.e., the Bidirectional Long Short-Term Memory (BiLSTM) network. Given sufficient training, the model is capable of achieving the automation of ionization edge recognition on raw EELS spectra. To build a large and diverse ground truth labeled training dataset, we digitized a large number of core-loss edges collected from real experiments and literature; we built the ground truth-labelled spectrum database by synthesizing spectra by including experimental conditions, including the multiple scattering effect, the uncertainty of the energy axis due to stray field and hysteresis of the prism, the low-frequency patterned noise, etc. Our well-trained CNN-BiLSTM network was tested against both the validation spectra and the real spectra collected from experiments, with average F-1 scores of 0.98 and 0.99, respectively. For the real spectra dataset prediction, only 1.1% of the predicted spectra have false-negative elements, and 94.9% of them are accurately predicted where the predicted edges exactly match the ground truth. The sensitivity of the network was further evaluated by detecting the lowest signal noise ratio at the edge peak (PSNR) for each element, at which the network can recognize the edge. Correspondingly, the detectable jump ratios were obtained. Based on our observation, the detectable jump ratios and PSNR are very low meaning that our networks are accurate in low-dose situations. It suggests that our network is remarkably sensitive to the edge signal. The high accuracy and sensitivity demonstrate that our network is highly reliable for core-loss edges recognition.

Methods

Database construction.   Considering that the available spectra from our own experiments are not enough for training a deep learning model, we built a spectrum database containing 250,000 simulated spectra, which are synthesized based on real ionization edges obtained from experiments/literature. The various physical phenomena in experiments are embedded into the synthesized spectra by convolving the broadening effects and simulating the instability with randomness. In addition, edges from different experiments can have different quality and energy resolutions, which help to further diversify the database. The simulated spectra are in close resemblance to the real EELS spectra and are suitable for our CNN-BiLSTM network training.

Edge modeling.   We digitized around 400 core-loss edges covering 20 different elements from journal papers. The edges are then fitted by using the multi-gaussian model to locate the coordinates of the peaks and smooth out the noise effects as shown in Fig. 1a[13]. Figure 1b presents some examples of the fitted edges for different elements. We can notice that the fitted curves round off the noise fluctuations well with the near edge fine features well preserved.

In the real case, when a spectrum is recorded by the detector, the instrument broadening is also convolved in it. Therefore, we further processed the modeled edges by convolving them with the corresponding broadening function, i.e., the point spread function (PSF). The intensity with scattering effect can be expressed as I_r:
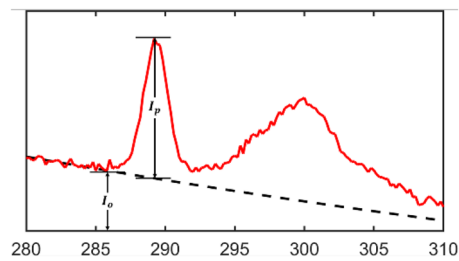
Figure 2. Jump ratio and peak signal noise ratio.

$$I_r(E) = I_u(E) \otimes PSF_{ZLP}$$

where I_u is the original intensity without the broadening effect and E is the energy loss.

The scattering effect is composed of amonochromaticity of the beam and broadening due to the detectors. Thus, PSF is constructed by two terms:

$$PSF_{ZLP} \propto PSF_{amono} \otimes PSF_{detector}$$

where PSF_{amono} represents the energy spread of the incoming electron beam, and $PSF_{detector}$ represents the broadening in the detectors. For the amonochromaticity effect, the form of the Gaussian distribution function is used:

$$PSF_{amono} \propto e^{-\frac{E^2}{2\sigma^2}}$$

where σ is the standard deviation, and can be expressed by:

$$\sigma = \frac{FWHM_g}{2\sqrt{2\ln 2}}$$

$FWHM_g$ is the full-width half-maximum of the Gaussian distribution. In our data augmentation, it ranges from 0.1 to 1.6.

For the detector broadening effect, the Lorentzian distribution function is commonly used to describe the diffused tail (caused by light diffusion in the scintillator and the optical fiber coupler), which is expressed by:

$$PSF_{detector} \propto \frac{HWHM_l}{E^2 + HWHM_l^2}$$

$HWHM_l$ is the full width at half maximum of the Lorentzian distribution. In our data augmentation, it ranges from 0.1 to 0.6.

With the broadening effects convolved, the synthesized edges will be comparable to those obtained from the different instruments. Figure 1c shows the Mn–L edge before and after convolving the PSF. We can observe that the broadening effect of the point spread functions (PSFs) has been successfully embedded into the modeled edge. It is worth mentioning that the Lorentzian broadening is often neglected by researchers in the field but it is a critical broadening effect that shall be included to reproduce an experimental EELS spectrum.

Background and jump ratio. In the core-loss region of the spectrum, it has been commonly accepted that the background can be modeled by using the inverse power law equation:

$$J(E) \propto E^{-r}$$

where r can vary across the specimen as a result of changes in thickness and composition. Typically, the value of r fluctuates around 1.88. In our case, we assign the value based on a normal distribution with a mean of 1.88 and a standard deviation of 0.2. The values less than 0.1 in the normal distribution are cut off to guarantee that the background remains realistic.

To embed the modeled edges into the background, the jump ratios of the edges must be assigned. The jump ratio is defined as the ratio of the intensity difference between the peak intensity and the background I_p and the intensity of the edge start point I_o, which is shown in Fig. 2. In our synthesized spectrum, the minimum jump ratio of the edges is set to 0.15. Because using data with too low a SNR for training would result in overfitting and high false positive rate, we choose a noticeable minimum jump ratio (> 0.15) to insure low false positive rate. Based on the digitized experimental data, we use the maximum jump ratios from experimental spectrum to set the upper limit of the jump ratios for the corresponding element. We summarized the edges of 20 elements and their edge types collected from the literature and the jump ratio range in Table 1. In our supplementary material, Table S1 lists the exact substances where these edges are collected.

Spectrum synthesize and data augmentation. The spectrum synthesis process is summarized in Fig. 3. The first step is to randomly pick 1 to 6 edges from the edge library and convolve them with the PSFs. Figure 3a shows

Element (edge type)	Jump ratio range	Number of edges	Element (edge type)	Jump ratio range	Number of edges
B (K)	0.15–0.5	31	Mn (L)	0.15–0.5	26
C (K)	0.15–1.5	25	Fe (L)	0.15–1.5	39
N (K)	0.15–0.7	14	Co (L)	0.15–0.7	23
O (K)	0.15–1.5	38	Ni (L)	0.15–1.5	20
F (K)	0.15–1.0	8	Cu (L)	0.15–1.0	17
Ca (L)	0.15–1.3	13	Si (L)	0.15–1.3	34
Sc (L)	0.15–1.3	6	P (L)	0.15–1.3	5
Ti (L)	0.15–1.5	47	S (L)	0.15–1.5	1
V (L)	0.15–1.5	40	Cl (L)	0.15–1.5	1
Cr (L)	0.15–1.5	18	Ar (L)	0.15–1.5	1

Table 1. Edge library[14–65].
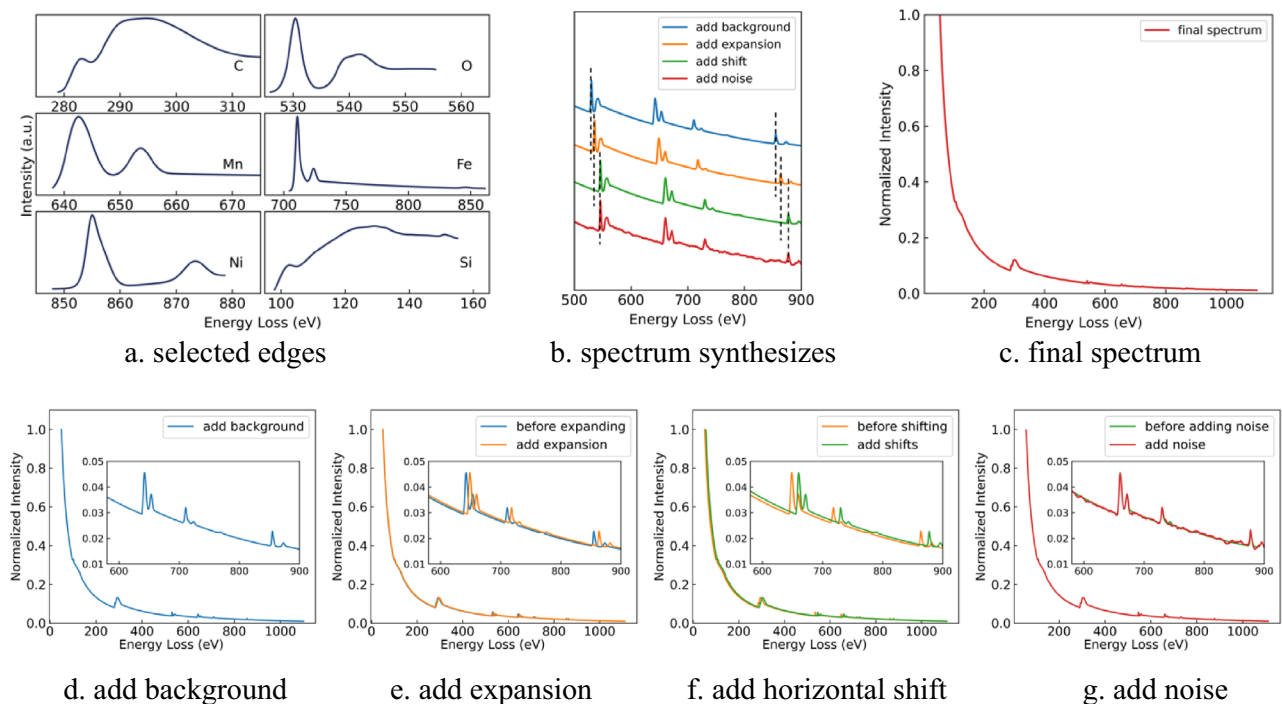
Figure 3. Spectrum synthesize steps and process.

the 6 convolved edges we used in the example. Figure 3b lists the augmentation operations before obtaining the final spectrum shown in Fig. 3c.

Since we only consider the core-loss region on spectra, the energy loss range is set from 50 to 1100 eV. The background $J(E)$ is normalized to $J_n(E)$:

$$J_n(E) = \frac{E^{-r}}{50^{-r}}$$

Before adding the edges to the background, to simulate the long tail of the core-loss edges and obtain a smooth merged spectrum, we extend the edges using another inverse power law, where the range of the power is from − 3.8 to − 1.8.

With the jump ratio randomly selected from the corresponding range for the element, the extended edge can be added to the background, which is shown in Fig. 3d.

In the real case, distortions and translational shifts of the spectra can be introduced due to multiple reasons including the hysteresis of the prism and the magnifying lenses, the spectrometer alignment, stray magnetic fields, and even temperature, etc. As a result, the spectra collected at different times can vary for the same instrument. To simulate these phenomena, we allocate a maximum of 1.5% of the contraction or expansion ratio to the modeled spectrum. A random shift of the spectrum is also implemented to the spectrum with a maximum shift of 15 eV. Figure 3b and e show the expanded spectrum. The spectrum that is further shifted is shown in Fig. 3b and f.
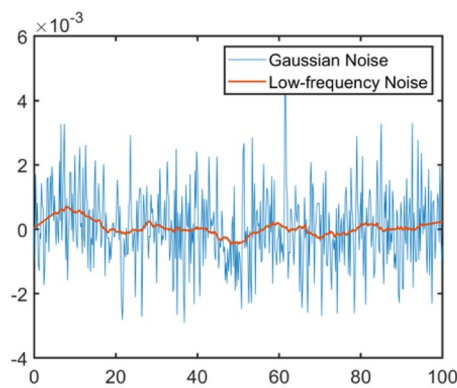
Figure 4. Gaussian noise and filtered noise.

Operation	Randomness range
Edge select	Types of elements: 1–6
PSFs convolve	$FWHM_g$: 0.1–1.6
	$HWHM_l$: 0.1–0.6
Edge extend	Power r': −3.8 to −1.8
Background	r: Gaussian (1.88, 0.2) > 0.1
Jump ratio	Range shown in Table 1
Contraction/expansion	−1.5 to 1.5%
Horizontal shift	−15 to 15 eV
Gaussian noise	Gaussian (0, 0.0012)
Savitzky–Golay filter	Frame length: 30–60 odd number
	Coefficient C: 1.0–2.0

Table 2. Augmentation operations summary.

The last step of the spectrum synthesis is to add correct noise to it. Firstly, Gaussian noise with a mean of 0 and a standard deviation of 0.0012 is generated. Afterward, the Savitzky–Golay filter is applied to smooth the noise to low-frequency noise, where the polynomial order of the filter is set to 1, and the frame length is an odd positive integer in the range of 30 to 60. Figure 4 shows the raw Gaussian noise and the filtered noise. The filtered low-frequency noise is also called fixed-pattern noise. It is frequently seen in experimental EELS spectra primarily due to the non-uniform gain on the scintillator. To augment the noise level, a random coefficient C that is in the range from 1 to 2 is further assigned to the noise. Figure 3g presents the spectrum with noise added, which is also the input spectrum for our deep learning network. Table 2 summarizes the data augmentation operations in the process of spectrum synthesis. We generated 250,000 spectra in total, where the percentage of the spectra containing 1 to 6 edges are 20%, 20%, 20%, 15%, 15%, and 10%, respectively.

Network construction. *CNN-BiLSTM network.* CNN is a type of network developed for conducting image recognition and has been widely implemented in image classification. A typical CNN structure is constructed by connecting a few convolutional layers and dense layers (fully connected layers). The convolutional layers can extract the features from the input data and the dense layers classify the features to achieve classification. The recognition of the ionization edges on a spectrum is also a type of image recognition task. In the meanwhile, the onset ionization edge carries the information of the target element. The edge's energy position relative to others plays a critical role in the classification process. However, a pure CNN structure is insensitive to the positional/sequential information. Given that the long short-term memory network (LSTM) enables the network to memorize the information in sequence by selectively keeping the important data and deleting less important data to pass to the next layer, we proposed a CNN-BiLSTM combined network for ionization edge recognition and classification in the core loss region of the spectrum. Figure 5b shows the architecture of our CNN-BiLSTM network, where the BiLSTM network is plotted in an unwrapped way. The real LSTM cell is described in Fig. 5c, and is constructed with 100 hidden units. In our case, the input layer carries the information of a spectrum that is a 1D vector (spectrum of energy loss range from 90 to 1000 eV with interval 0.2 eV). A filter of size 11 is implemented in each convolutional layer. Batch normalization operation[66] is carried out following each convolutional layer, and 20% of node dropout regularization is utilized after each dense layer to avoid overfitting. Activation functions, ReLU and LeakyReLU of scale 0.01, are implemented to connect different layers as shown in the Fig. 5b. The network is built and trained using MATLAB. The training parameter settings are listed
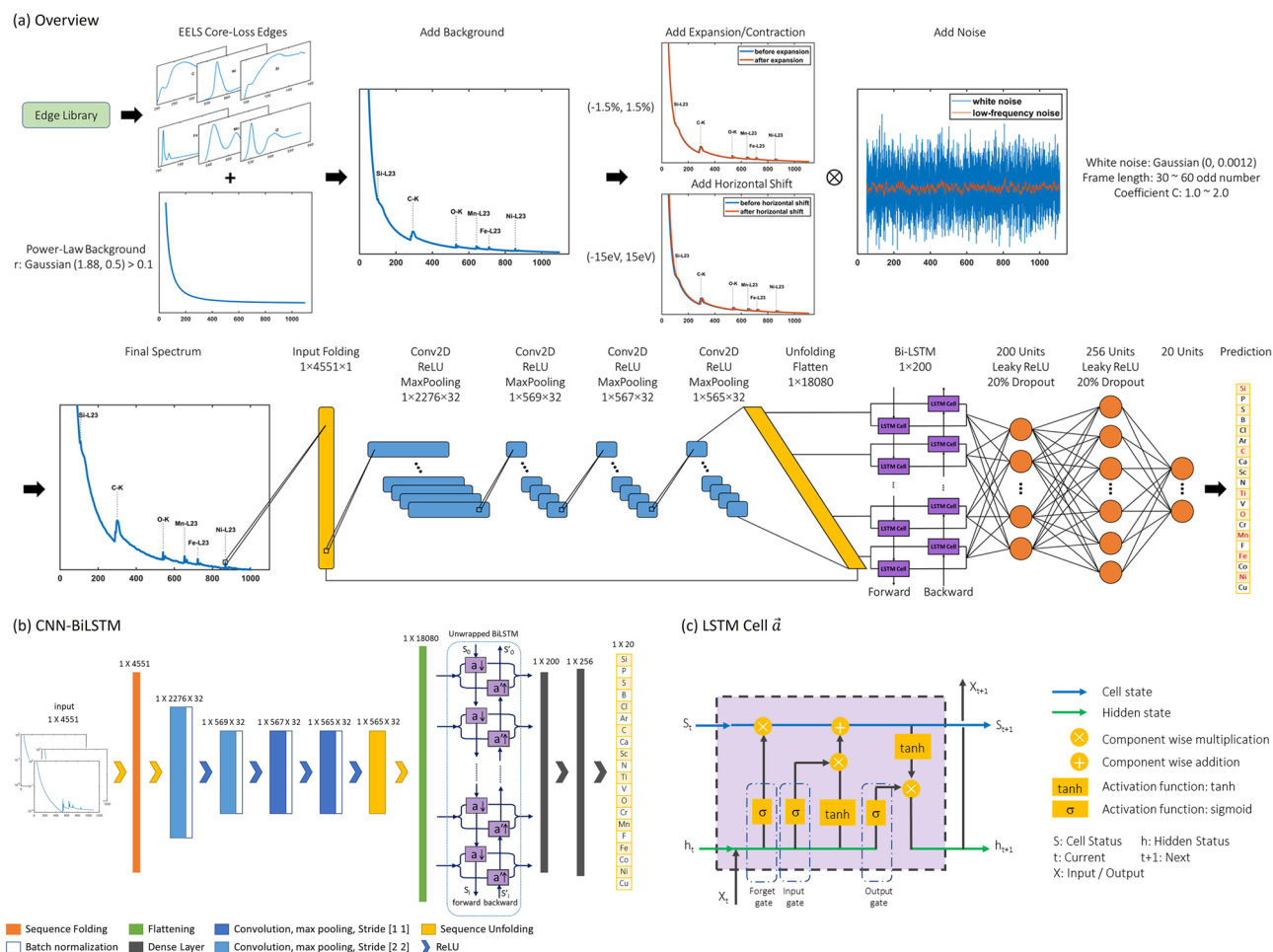
Figure 5. CNN-BiLSTM network architecture. (**a**) The overview. (**b**) The CNN-BiLSTM network structure. (**c**) The structure of the LSTM cell.

Options	Setting
ExecutionEnvironment	GPU
InitialLearnRate	0.001
GradientThreshold	1
Shuffle	Every-epoch
ValidationPatience	Inf
ValidationFrequency	50
Verbose	True

Table 3. Network training setting.

in Table 3. 80% of the spectra in the database are randomly picked for the training set, and the remaining 20% are for the validation set.

Results and discussion

Prediction validation. The performance of the proposed CNN-BiLSTM network is validated using multiple metrics.

F1-score. To evaluate the performance of the network per spectrum, we calculate the F1-score for each predicted spectrum, which is defined as the harmonic mean of precision and recall. We use NE_c, NE_p, and NE_r to represent the number of correctly predicted edges (true positive), the total number of predicted edges (predicted

6

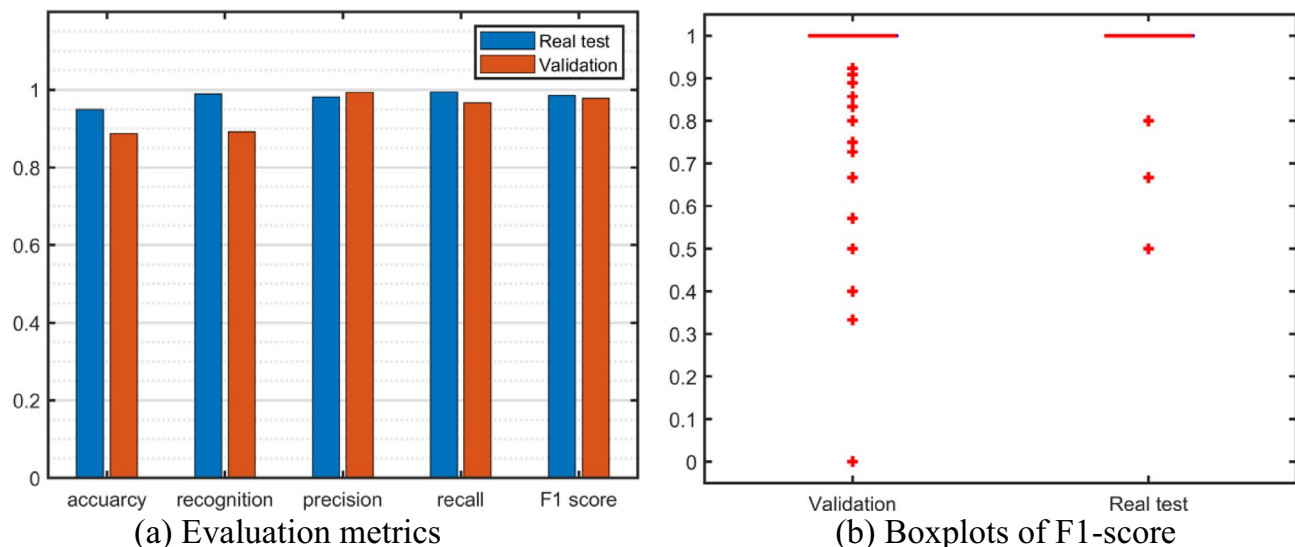(a) Evaluation metrics   (b) Boxplots of F1-score

Figure 6. Performance of the well-trained CNN-BiLSTM network.

positive), and the number of real edges (actual positive) on the spectrum, respectively. Precision, recall, and F1-score can be calculated by using the following equations:

$$precision = \frac{NE_c}{NE_p}$$

$$recall = \frac{NE_c}{NE_r}$$

$$F1 = \frac{2}{recall^{-1} + precision^{-1}}$$

Exact accuracy. We use exact accuracy (ACC) to check how our network can accurately recognize all the edges on a spectrum, i.e., the ratio of the number of spectra accurately predicted (N_{exa}) to the total number of spectra tested (N). This value also reflects the percentage of the spectra whose F1-score equals to 1.

$$Acc = \frac{N_{exa}}{N}$$

Recognition accuracy. The following equation shows the format of the recognition accuracy. N_{rec} represents the number of spectra that the network can recognize all the edges that appear on them. In other words, the percentage of the spectra whose recall is 1. This matric suggests how sensitive the network can detect the available edges.

$$Rec = \frac{N_{rec}}{N}$$

Network validation. Besides the synthesized validation dataset, we also use 177 real experimental spectra to examine the performance of our network. The real spectra are from EELS Atlas Database from EELS.info, EELS DB (EELS Data Base) and our own experimental data[67]. Detailed experimental conditions for each spectra, including spectrum acquisition instrument, energy resolution, etc. are listed in Table S2 of the supplementary material.

Figure 6a shows the values of validation metrics for both the real spectra dataset and the validation dataset (The precision, recall, and F1-score are the averaged values). Figure 6b presents the boxplots of the F1-scores of the two datasets, where only some outliers have F1-scores less than 1, which means that our network is highly accurate in recognizing the edges. The average F1-scores of the real spectra dataset and the validation dataset are as high as 0.99 and 0.98, respectively. The comparable high F1-scores of the two datasets further indicate that the network is well-trained without overfitting. The average precision of the real spectra dataset and the validation dataset are 0.98 and 0.99, respectively. The average recall of the real spectra dataset and the validation dataset are 0.99 and 0.97, respectively. The high values of the precision and recall indicate that false positive and false negative edges barely appear in a prediction.

The exact accuracy and the recognition accuracy are two metrics describing the percentages of the desired spectra. For the real dataset, the exact accuracy reaches 94.9%. And the network sensitively detects all the edges presented on the spectra, which can be reflected by the recognition accuracy of 98.9%. In other words, only
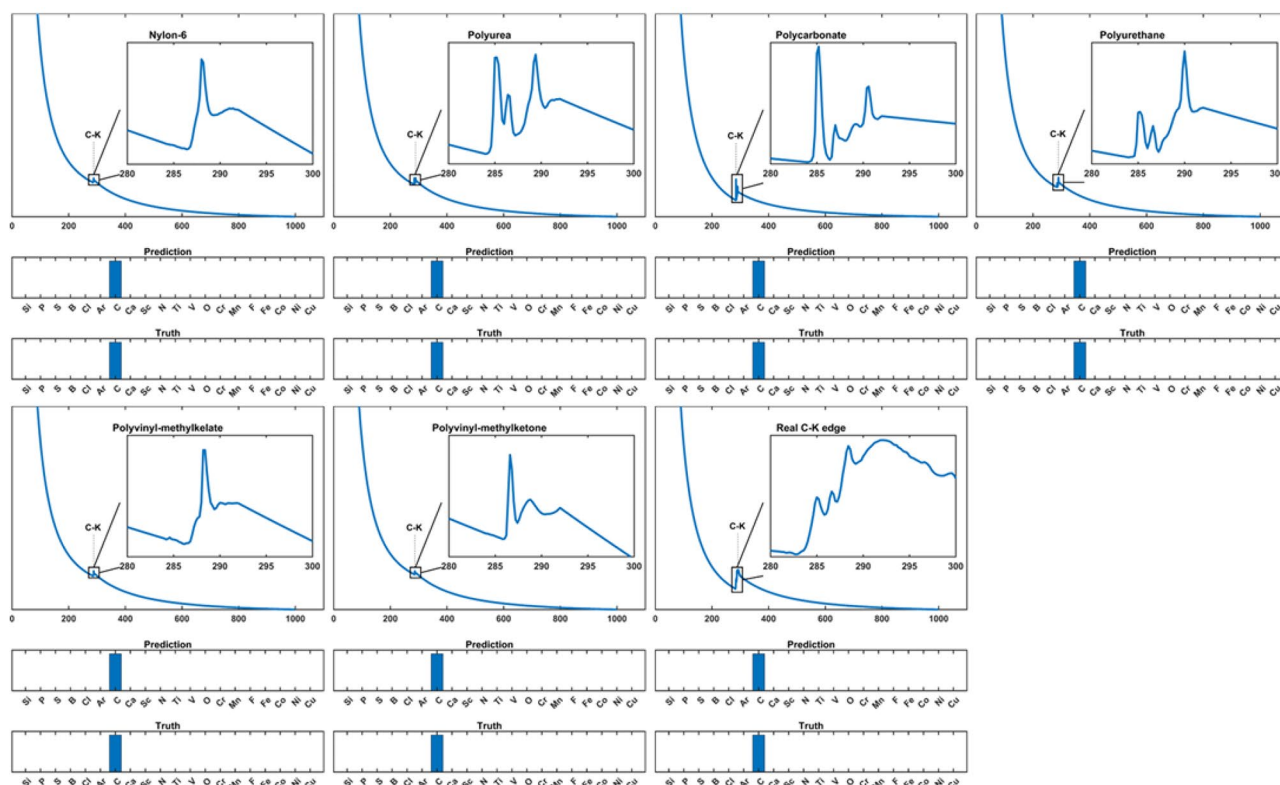
Figure 7. Network prediction versus ground truth for C-K edges from different materials.

around 1% of the predicted spectra have edges that failed to be detected by our network. Indeed, among the 9 failed cases, almost all the false positive Edges are due to the misrecognition of the nitrogen edge and scandium edge. Because the edges of these two elements have similar shapes and onsets, they can look identical on a spectrum. The false negative edges appear when the jump ratio is too low to be visible. The exact accuracy and recognition accuracy of the validation set is relatively lower than those of the real spectra dataset, but they are still around 90%. This phenomenon could be caused by the higher diversity of the synthesized spectra, where we stretched the parameters' ranges during data augmentation to create more extreme cases.

Considering that the same element presenting in different compounds or allotropes can generate edges with various fine structures, the capability of our network in recognizing these edges is further evaluated. Carbon, as one of the essential elements forming the most compounds on earth and has multiple allotropes, is chosen as an example in this section to test our network. Figure 7 shows the C-K edge prediction of our network compared with the ground truth. Our network precisely detected the existence of the C-K edges in different materials containing carbon element. In addition, those spectra are from experiments[68,69] different from those in our edge library.

Evaluation of model's sensitivity.   *Peak signal noise ratio (PSNR).*   To further evaluate the sensitivity of the model in detecting the ionization edges for a specific element, we introduce the concept of Peak Signal Noise Ratio, which is the signal noise ratio at the peak of the edge. It is expressed by the ratio of the signal at the edge peak off the inverse power law background (shown in Fig. 2 as I_p) to the standard deviation of the noise (*sd*). In other words,

$$PSNR = \frac{I_p}{sd}$$

To make the value of PSNR for different elements comparable, we further synthesize one spectrum for each element with the value of the power *r* for the background set to be the same constant 1.88. We added different levels of the Gaussian noise (white noise) to the spectra to create more extreme situations to examine the sensitivity of our network. The standard deviation of the noise is set to be one-tenth of the minimum value of the background:

$$std = \frac{\min_{50\sim1100} J_n(E)}{10}$$

Element (edge type)	PSNR	Jump ratio	Element (edge type)	PSNR	Jump ratio
B (K)	23.00	**0.077**	Mn (L)	3.50	**0.090**
C (K)	10.40	**0.071**	Fe (L)	3.68	**0.116**
N (K)	5.87	**0.068**	Co (L)	3.52	**0.130**
O (K)	4.51	**0.088**	Ni (L)	3.05	**0.126**
F (K)	3.45	**0.099**	Cu (L)	3.16	**0.158**
Ca (L)	7.69	**0.075**	Si (L)	56.29	**0.059**
Sc (L)	6.29	**0.078**	P (L)	50.26	**0.086**
Ti (L)	5.36	**0.083**	S (L)	19.26	**0.049**
V (L)	3.99	**0.070**	Cl (L)	16.82	**0.062**
Cr (L)	3.85	**0.083**	Ar (L)	7.94	**0.040**

Table 4. Detectable PSNR and jump ratio. Significant values are in bold.

Sensitivity evaluation.   To quantify the sensitivity of our network, we searched for the minimum value of *PSNR* for each element that the network can correctly recognize by varying the scale of the peak. To avoid anomalous results, for each element, we test the minimum value of *PSNR* for each edge in the library for 100 times and use the arithmetic mean as the final detectable *PSNR*. Correspondingly, the value of the jump ratio can be captured simultaneously. The results are listed in Table 4.

It is worth noting that the detectable jump ratios in Table 4 are much lower than the minimum value of 0.15 we set for building the training dataset. This result illustrates that our network can manage the situation when the jump ratios of the edges are lower than those commonly observed in experiments. In the meanwhile, the noise is not smoothed by the Savitzky–Golay filter at this stage. Still, our network can sensitively extract valuable information from the noisy spectra. Thus, the detectable *PSNR* and Jump ratio in Table 4 can also serve as a reference for researchers determining whether the edges obtained from their experiments can be recognized by our network. Higher detectable *PSNRs* for light elements are expected because these edges tend to appear in a relatively high-intensity lower energy loss region. Referring to the corresponding low jump ratios, the network performs well in recognizing these light elements' ionization edges.

Conclusion

In this work, we built a CNN-BiLSTM network for automatically detecting the ionization edges on EELS spectra. We also created a forward model for synthesizing EELS spectra, which allows us to create a database containing 250,000 spectra with ground-truth labels for training and testing the network. The network is examined by using both the synthesized validation dataset and the real spectra dataset collected from experiments. Multiple validation metrics are calculated to evaluate the performance of the network regarding the two datasets, including exact accuracy, recognition accuracy, average precision, average recall, and average F1-score. The sensitivity of the network is explored by testing the lowest detectable PSNR and detectable jump ratio.

The average F1-score, recall, and precision for the real spectra dataset agree well with those for the validation dataset, which suggests that, for one thing, the synthesized spectra are comparable to real spectra, for the other, the network is well-trained without overfitting. The high values of these three metrics also illustrate the good performance of the network in respect to the low false positive or false negative rate. We find that for 98.9% of the real spectra, our network is able to recognize all the edges. For 94.9% of them, our network can accurately recognize all the edges without false positive edges, which means that only around 4% of the spectra have false positive edges, and around 1% of the spectra have false negative edges. What's more, the low value of the detectable PSNRs and jump ratios further demonstrate the high sensitivity of our network.

Taken together, our CNN-BiLSTM network can automatically recognize the core-loss edges on a raw spectrum with high accuracy. It is also safe to be applied to EELS spectra where the SNR and jump ratio are extremely low. The success of the network in automatically recognizing core-loss edges on a raw spectrum further makes it possible to the automation of more advanced core-loss edge analysis.

Data availability

Data and code are available from the corresponding author upon request.

References

1. Muller, D. A. Structure and bonding at the atomic scale by scanning transmission electron microscopy. *Nat. Mater.* **8**, 263–270 (2009).
2. Keast, V. J. Application of EELS in materials science. *Mater. Charact.* **73**, 1–7 (2012).
3. Gao, Z., Li, A., Ma, D. & Zhou, W. Electron energy loss spectroscopy for single atom catalysis. *Top. Catal.* https://doi.org/10.1007/s11244-022-01577-7 (2022).
4. Verbeeck, J. & Van Aert, S. Model based quantification of EELS spectra. *Ultramicroscopy* **101**, 207–224 (2004).

5. Bonnet, N., Brun, N. & Colliex, C. Extracting information from sequences of spatially resolved EELS spectra using multivariate statistical analysis. *Ultramicroscopy* **77**, 97–112 (1999).
6. Bonnet, N. & Nuzillard, D. Independent component analysis: A new possibility for analysing series of electron energy loss spectra. *Ultramicroscopy* **102**, 327–337 (2005).
7. de la Peña, F. *et al.* Mapping titanium and tin oxide phases using EELS: An application of independent component analysis. *Ultramicroscopy* **111**, 169–176 (2011).
8. Cueva, P., Hovden, R., Mundy, J. A., Xin, H. L. & Muller, D. A. Data processing for atomic resolution electron energy loss spectroscopy. *Microsc. Microanal.* **18**, 667–675 (2012).
9. Carey, C., Boucher, T., Mahadevan, S., Bartholomew, P. & Dyar, M. D. Machine learning tools for mineral recognition and classification from Raman spectroscopy. *J. Raman Spectrosc.* **46**, 894–903 (2015).
10. Pate, C. M., Hart, J. L. & Taheri, M. L. RapidEELS: Machine learning for denoising and classification in rapid acquisition electron energy loss spectroscopy. *Sci. Rep.* **11**, 19515 (2021).
11. Chatzidakis, M. & Botton, G. A. Towards calibration-invariant spectroscopy using deep learning. *Sci. Rep.* **9**, 2126 (2019).
12. Fung, K. L. Y. *et al.* Accurate EELS background subtraction—An adaptable method in MATLAB. *Ultramicroscopy* **217**, 113052 (2020).
13. Ahn, C. C. *Transmission Electron Energy Loss Spectrometry in Materials Science and The EELS Atlas* (Wiley, 2004). https://doi.org/10.1002/3527605495.
14. Garvie, L. A. J., Craven, A. J. & Brydson, R. Parallel electron energy-loss spectroscopy (PEELS) study of B in minerals; The electron energy-loss near-edge structure (ELNES) of the B K edge. *Am. Mineral.* **80**, 1132–1144 (1995).
15. Liao, Y. *Practical Electron Microscopy and Database* (2006).
16. Xie, K. Y. *et al.* Microstructural characterization of boron-rich boron carbide. *Acta Mater.* **136**, 202–214 (2017).
17. Gilbert, B. *et al.* Multiple scattering calculations of bonding and X-ray absorption spectroscopy of manganese oxides. *J. Phys. Chem. A* **107**, 2839–2847 (2003).
18. Carroll, K. J. *et al.* Probing the electrode/electrolyte interface in the lithium excess layered oxide $Li_{1.2}Ni_{0.2}Mn_{0.6}O_2$. *Phys. Chem. Chem. Phys.* **15**, 11128 (2013).
19. Lin, F. *et al.* Influence of synthesis conditions on the surface passivation and electrochemical behavior of layered cathode materials. *J Mater Chem A* **2**, 19833–19840 (2014).
20. Tan, H., Verbeeck, J., Abakumov, A. & Van Tendeloo, G. Oxidation state and chemical shift investigation in transition metal oxides by EELS. *Ultramicroscopy* **116**, 24–33 (2012).
21. Fernández, A. *et al.* Characterization of carbon nitride thin films prepared by dual ion beam sputtering. *Appl. Phys. Lett.* **69**, 764–766 (1996).
22. Garvie, L. A. J. & Buseck, P. R. Prebiotic carbon in clays from Orgueil and Ivuna (CI), and Tagish Lake (C2 ungrouped) meteorites. *Meteorit. Planet. Sci.* **42**, 2111–2117 (2007).
23. Feng, Z., Lin, Y., Tian, C., Hu, H. & Su, D. Combined study of the ground and excited states in the transformation of nanodiamonds into carbon onions by electron energy-loss spectroscopy. *Sci. Rep.* **9**, 3784 (2019).
24. Cavé, L., Al, T., Loomer, D., Cogswell, S. & Weaver, L. A STEM/EELS method for mapping iron valence ratios in oxide minerals. *Micron* **37**, 301–309 (2006).
25. Pool, V. *et al.* Site determination and magnetism of Mn doping in protein encapsulated iron oxide nanoparticles. *J. Appl. Phys.* **107**, 09B517 (2010).
26. Knappett, B. R. *et al.* Characterisation of Co@Fe3O4 core@shell nanoparticles using advanced electron microscopy. *Nanoscale* **5**, 5765 (2013).
27. Dash, S. S., Mukherjee, P., Haskel, D., Rosenberg, R. A. & Levy, M. Boosting optical nonreciprocity: Surface reconstruction in iron garnets. *Optica* **7**, 1038 (2020).
28. Wang, C. *et al.* Structure versus properties in α-Fe_2O_3 nanowires and nanoblades. *Nanotechnology* **27**, 035702 (2016).
29. Brück, S. *et al.* Magnetic and electronic properties of the interface between half metallic Fe_3O_4 and semiconducting ZnO. *Appl. Phys. Lett.* **100**, 081603 (2012).
30. Jin, Y., Xu, H. & Datye, A. K. Electron energy loss spectroscopy (EELS) of Iron Fischer–Tropsch catalysts. *Microsc. Microanal.* **12**, 124–134 (2006).
31. Chang, Y. K. *et al.* X-ray absorption of Si–C–N thin films: A comparison between crystalline and amorphous phases. *J. Appl. Phys.* **86**, 5609–5613 (1999).
32. Córdoba, R. *et al.* Nanoscale chemical and structural study of Co-based FEBID structures by STEM-EELS and HRTEM. *Nanoscale Res. Lett.* **6**, 592 (2011).
33. Meng, M. *et al.* Three dimensional band-filling control of complex oxides triggered by interfacial electron transfer. *Nat. Commun.* **12**, 2447 (2021).
34. Tyunina, M. *et al.* Oxygen vacancy dipoles in strained epitaxial $BaTiO_3$ films. *Phys. Rev. Res.* **2**, 023056 (2020).
35. Liu, H. *et al.* Spatially resolved surface valence gradient and structural transformation of lithium transition metal oxides in lithium-ion batteries. *Phys. Chem. Chem. Phys.* **18**, 29064–29075 (2016).
36. Lin, F. *et al.* Profiling the nanoscale gradient in stoichiometric layered cathode particles for lithium-ion batteries. *Energy Environ. Sci.* **7**, 3077 (2014).
37. Saitoh, M. *et al.* Systematic analysis of electron energy-loss near-edge structures in Li-ion battery materials. *Phys. Chem. Chem. Phys.* **20**, 25052–25061 (2018).
38. Yang, Y. *et al.* In situ TEM observation of resistance switching in titanate based device. *Sci. Rep.* **4**, 3890 (2015).
39. Niu, G. *et al.* On the local electronic and atomic structure of $Ce_{1-x}Pr_xO_{2-\delta}$ epitaxial films on Si. *J. Appl. Phys.* **116**, 123515 (2014).
40. Tomita, K., Miyata, T., Olovsson, W. & Mizoguchi, T. Strong excitonic interactions in the oxygen K-edge of perovskite oxides. *Ultramicroscopy* **178**, 105–111 (2017).
41. Wang, L. *et al.* Electrochemically driven giant resistive switching in perovskite nickelates heterostructures. *Adv. Electron. Mater.* **3**, 1700321 (2017).
42. Wang, R., Jiang, W., Xia, D., Liu, T. & Gan, L. Improving the wettability of thin-film rotating disk electrodes for reliable activity evaluation of oxygen electrocatalysts by triggering oxygen reduction at the catalyst-electrolyte-bubble triple phase boundaries. *J. Electrochem. Soc.* **165**, F436–F440 (2018).
43. Hirayama, K., Ii, S. & Tsurekawa, S. Transmission electron microscopy/electron energy loss spectroscopy measurements and ab initio calculation of local magnetic moments at nickel grain boundaries. *Sci. Technol. Adv. Mater.* **15**, 015005 (2014).
44. Vilá, R. A., Huang, W. & Cui, Y. Nickel impurities in the solid-electrolyte interphase of lithium-metal anodes revealed by cryogenic electron microscopy. *Cell Rep. Phys. Sci.* **1**, 100188 (2020).
45. Sina, M. *et al.* Structural phase transformation and Fe valence evolution in FeO_xF_{2-x}/C nanocomposite electrodes during lithiation and de-lithiation processes. *J. Mater. Chem. A* **1**, 11629 (2013).
46. Laffont, L. *et al.* High resolution EELS of Cu–V oxides: Application to batteries materials. *Micron* **37**, 459–464 (2006).
47. Wang, C. *et al.* Novel hybrid nanocomposites of polyhedral Cu_2O nanoparticles–CuO nanowires with enhanced photoactivity. *Phys. Chem. Chem. Phys.* **16**, 17487–17492 (2014).
48. Li, W. & Ni, C. Electron energy loss spectroscopy (EELS). In *Encyclopedia of Tribology* (eds Wang, Q. J. & Chung, Y.-W.) 940–945 (Springer, US, 2013). https://doi.org/10.1007/978-0-387-92897-5_1223.

49. Wu, L. *et al.* Enhanced thermoelectric performance in Cu-intercalated BiTeI by compensation weakening induced mobility improvement. *Sci. Rep.* **5**, 14319 (2015).
50. Golla-Schindler, U., Benner, G., Orchowski, A. & Kaiser, U. In situ observation of electron beam-induced phase transformation of CaCO₃ to CaO via ELNES at low electron beam energies. *Microsc. Microanal.* **20**, 715–722 (2014).
51. Rossi, A. L. *et al.* Effect of strontium ranelate on bone mineral: Analysis of nanoscale compositional changes. *Micron* **56**, 29–36 (2014).
52. Yamada, M. *et al.* Reaction mechanism of "SiO"–carbon composite-negative electrode for high-capacity lithium-ion batteries. *J. Electrochem. Soc.* **159**, A1630–A1635 (2012).
53. Ma, J. W. *et al.* Carrier mobility enhancement of tensile strained Si and SiGe nanowires via surface defect engineering. *Nano Lett.* **15**, 7204–7210 (2015).
54. Homma, K., Kambara, M. & Yoshida, T. High throughput production of nanocomposite SiO_x powders by plasma spray physical vapor deposition for negative electrode of lithium ion batteries. *Sci. Technol. Adv. Mater.* **15**, 025006 (2014).
55. Sato, Y. K., Kuwauchi, Y., Miyoshi, W. & Jinnai, H. Visualization of chemical bonding in a silica-filled rubber nanocomposite using STEM-EELS. *Sci. Rep.* **10**, 21558 (2020).
56. Song, M., Fukuda, Y. & Furuya, K. Local chemical states and microstructure of photoluminescent porous silicon studied by means of EELS and TEM. *Micron* **31**, 429–434 (2000).
57. Jia, Y. *et al.* The effects of oxygen in spinel oxide $Li_{1+x}Ti_{2-x}O_{4-\delta}$ thin films. *Sci. Rep.* **8**, 3995 (2018).
58. Gloter, A., Ewels, C., Umek, P., Arcon, D. & Colliex, C. Electronic structure of titania-based nanotubes investigated by EELS spectroscopy. *Phys. Rev. B* **80**, 035413 (2009).
59. Huang, C.-N. *et al.* Nonstoichiometric titanium oxides via pulsed laser ablation in water. *Nanoscale Res. Lett.* **5**, 972–985 (2010).
60. Kitta, M., Akita, T., Tanaka, S. & Kohyama, M. Characterization of two phase distribution in electrochemically-lithiated spinel $Li_4Ti_5O_{12}$ secondary particles by electron energy-loss spectroscopy. *J. Power Sources* **237**, 26–32 (2013).
61. Won, S., Lee, S. Y., Hwang, J., Park, J. & Seo, H. Electric field-triggered metal-insulator transition resistive switching of bilayered multiphasic VO_x. *Electron. Mater. Lett.* **14**, 14–22 (2018).
62. Sigle, W. Analytical transmission electron microscopy. *Annu. Rev. Mater. Res.* **35**, 239–314 (2005).
63. Chi, M. *et al.* Atomic and electronic structures of the $SrVO_3$-$LaAlO_3$ interface. *J. Appl. Phys.* **110**, 046104 (2011).
64. Lin, H.-T., Nayak, P. K., Wang, S.-C., Chang, S.-Y. & Huang, J.-L. Electron-energy loss spectroscopy and Raman studies of nanosized chromium carbide synthesized during carbothermal reduction process from precursor $Cr(CO)_6$. *J. Eur. Ceram. Soc.* **31**, 2481–2487 (2011).
65. Bindi, L. *et al.* Kishonite, VH2, and oreillyite, Cr2N, two new minerals from the corundum xenocrysts of Mt. Carmel, Northern Israel. *Minerals* **10**, 1118 (2020).
66. Ioffe, S. & Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. Preprint at http://arxiv.org/abs/1502.03167 (2015).
67. Ewels, P., Sikora, T., Serin, V., Ewels, C. P. & Lajaunie, L. A complete overhaul of the electron energy-loss spectroscopy and X-ray absorption spectroscopy database: eelsdb. eu. *Microsc. Microanal.* **22**, 717–724 (2016).
68. Alexander, D., Anderson, J., Forró, L. & Crozier, P. The real carbon K-edge. *Microsc. Microanal.* **14**, 674–675 (2008).
69. Ade, H. & Stoll, H. Near-edge X-ray absorption fine-structure microscopy of organic and magnetic materials. *Nat. Mater.* **8**, 281–290 (2009).

Acknowledgements

Author contributions

H.L.X. conceived the idea. Z.J. and L.K. performed the experiments. L.K., Z.J. and H.L.X. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-25870-3.

Correspondence and requests for materials should be addressed to H.L.X.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.