LETTER • OPEN ACCESS

Deep learning of interface structures from simulated 4D STEM data: cation intermixing vs. roughening*

MACHINE LEARNING
Science and Technology

Deep learning of interface structures from simulated 4D STEM data: cation intermixing vs. roughening[*]

M P Oxley[1], J Yin[2], N Borodinov[1], S Somnath[2], M Ziatdinov[1,3], A R Lupini[1], S Jesse[1], R K Vasudevan[1] and S V Kalinin[1]

1   The Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, TN 37831, United States of America
2   National Center for Computational Sciences, Oak Ridge National Laboratory, Oak Ridge, TN 37831, United States of America
3   Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, United States of America

E-mail: vasudevanrk@ornl.gov and sergei2@ornl.gov

Abstract

Interface structures in complex oxides remain an active area of condensed matter physics research, largely enabled by recent advances in scanning transmission electron microscopy (STEM). Yet the nature of the STEM contrast in which the structure is projected along the given direction precludes separation of possible structural models. Here, we utilize deep convolutional neural networks (DCNN) trained on simulated 4D STEM datasets to predict structural descriptors of interfaces. We focus on the widely studied interface between $LaAlO_3$ and $SrTiO_3$, using dynamical diffraction theory and leveraging high performance computing to simulate thousands of possible 4D STEM datasets to train the DCNN to learn properties of the underlying structures on which the simulations are based. We test the DCNN on simulated data and show that it is possible (with >95% accuracy) to identify a physically rough from a chemically diffuse interface and create a DCNN regression model to predict step positions. We quantify the applicability of the model to different thicknesses and the transferability of the approach. The method shown here is general and can be applied for any inverse imaging problem where forward models are present.

Oxide interfaces have emerged as one of the central topics in condensed matter physics research due to the multitude of unique behaviors they exhibit, ranging from interfacial conduction between dielectric materials [1, 2], superconductivity, magnetic effects [3–5], improper ferroelectric responses [6], and large ionic conductivity [7]. This breadth of functional responses has stimulated outstanding theoretical efforts aimed at understanding the corresponding functionalities and their relationship with the interface structure. These discoveries have been enabled by the advances in oxide growth techniques including pulsed laser deposition, magnetron sputtering, and molecular beam epitaxy, often augmented by the reflection high energy electron diffraction as a control and monitoring tool that allows for the sub-unit cell precision.

A number of functional mechanisms for emergence of novel physical behaviors at the interfaces have been identified, ranging from classical semiconducting behaviors associated with mismatch between band offsets and resultant band bending, charge injection and changes in oxidation states, to symmetry mismatch across the interface and corresponding penetration of the distortions associated with zone center (e.g. ferroelastic and polarization) or zone boundary (e.g. octahedra tilts) [8–10] modes. For instance, a broad set of phenomena has been found to emerge at ferroelectric interfaces both due to the polarization screening and lattice symmetry breaking [11]. Combined with the field-induced switching, this resulted in significant interest to these material systems as potential multiferroics. However, the physical effects at the interface can

compete with chemical interactions such as oxygen vacancy redistribution [12, 13], compensating for the band offset effects [14] and polarization screening [15]. Many of these phenomena were explored and understood via recent advances in aberration-corrected scanning transmission electron microscopy (STEM), where the combination of the direct observations of atomic position enabled the reconstruction of order parameter fields and chemical strains, whereas electron energy loss spectroscopy (EELS) allowed insight into local charge states of cations and oxygen stoichiometry.

However, the fundamental limitation of STEM is the fact that the structural and chemical information is averaged along the beam path, limiting the separability of dissimilar interfacial mechanisms. One of the outstanding issues in this area is the potential cation mixing at the interfaces [16], i.e. exchange between iso- or aliovalent cations in the adjacent layers during the growth. As can be readily understood, intermixing is expected to lead to effective doping in the adjacent layers, forming conductive channels, introducing new phases, etc. At the same time, in the projected image, intermixing will be indistinguishable from interface roughness, e.g. presence of substrate steps running across the interface, or island growth. Comparison of simulated and experimental EELS profiles have been used to determine if similar interfaces are abrupt [17]. However, if the experimental EELS profile across the interface is broader than expected, there is insufficient sensitivity in the EELS signal to differentiate between diffusion and interface roughness.

Here, we theoretically explore whether the cation intermixing and interface roughening can be distinguished using 4D STEM. In this technique, the 2D diffraction pattern from the sub-atomically focused electron beam is detected at each spatial location, giving rise to an information-rich 4D data set. However, the local diffraction pattern is determined both by the local material structure and the beam parameters including residual aberrations, spatial and temporal incoherence, obviating direct inversion. Here, we explore whether deep learning methods can be used to distinguish interface mechanisms from simulated 4D STEM data.

The main methodology used here is applicable to many inverse problems in imaging and is highlighted in figure 1. The range of physically realizable models of the material system corresponding to different classes of behavior or within a (small dimensional) parameter space are created, and their corresponding experimental fingerprints are modeled using a forward model. The corresponding model parameters or class are used as structural descriptors. A deep convolutional neural network (DCNN) is trained using the experimental fingerprints as an input and the structural descriptors as an output. Once trained on the corresponding training set, the network can be used on the unknown data to identify relevant structural descriptors. Notably, a similar approach was implemented using back-propagation neural networks for spectral data [18, 19] however, poor generalizability and the ad-hoc approach to the formation of feature vectors precluded broad applicability. More recent works have utilized DCNNs for similar inverse problems in imaging, for instance see [20–22]. The advantage of the convolutional operation being that it leverages shift invariance and weight sharing: the learned filters can be useful in multiple different parts of the same image, and by only passing information from small regions to subsequent convolutional layers, the dimensionality is greatly reduced compared with fully connected networks [23].

As a model system, we explored the most widely studied oxide interface in the recent past: i.e. the interface between the two band insulators $LaAlO_3$ (LAO) and $SrTiO_3$ (STO) [1, 24, 25]. We began with structural models as indicated in figures 2(a)–(c). We model the system via a unit-cell-wide cross-section across the LAO-STO interface as shown in figure 2(a), with varying thicknesses in the [001] direction (typical of growth on (001) $SrTiO_3$ substrates). The STO substrate is terminated with a TiO_2 layer, while the last LAO layer is terminated with an AlO layer. For maintaining periodic boundary conditions, a single SrO layer is added to cap the AlO. The section in figure 2(a) is what would be directly imaged by the electron beam. However, numerous structural possibilities exist in the direction of the beam, i.e. the [001] direction, including the existence of a buried step which causes roughening (as shown in figure 2(b)) and the B-site diffusion, i.e. of Ti into the Al sites. We therefore constructed dozens of possible structures where we varied the position of the buried interface, or alternatively, for a flat interface we varied the percentage of Ti substituting the Al in the first AlO layer at the interface (figure 2(c)). We then repeated these structures for varying thicknesses, from 100 Å to 300 Å. Once given the structures, we used the μSTEM simulation code [26, 27], to compute the convergent beam electron diffraction (CBED) patterns from all the generated structures. CBEDs are highly sensitive to local electric fields and therefore atomic-scale structural features [28]. The quantum excitation of phonons algorithm was used. Each CBED dataset was of size $(341 \times 341 \times 13 \times 123)$ In total the simulation size was on the order of ~350 GB. The multislice routine was carried out using slices half the dimensions of the basic STO structure, with the total number of slices dependent on the simulated thickness. The structure was tiled eight times in the y-direction to for the supercell and 8 configurations were calculated for each slice. This corresponds to a total of 64 configurations when the quick shifting option in μSTEM is used. Fifty passes were used at each probe position.
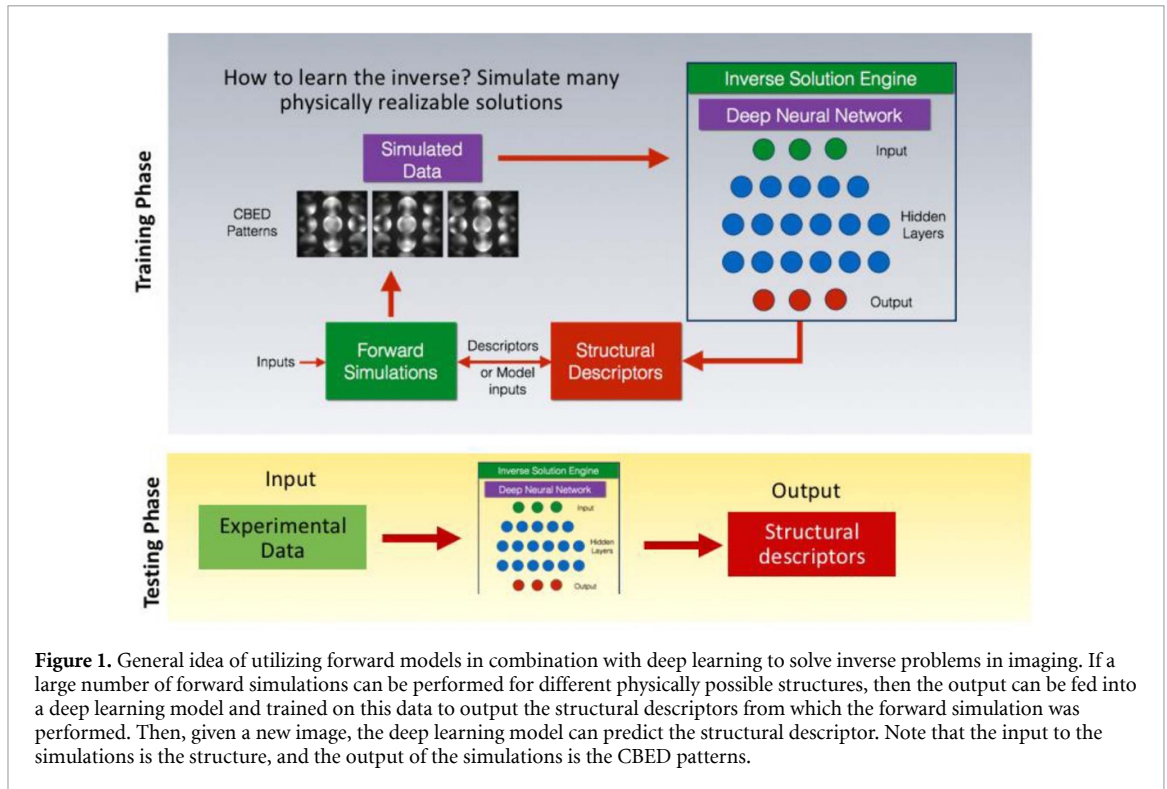
Figure 1. General idea of utilizing forward models in combination with deep learning to solve inverse problems in imaging. If a large number of forward simulations can be performed for different physically possible structures, then the output can be fed into a deep learning model and trained on this data to output the structural descriptors from which the forward simulation was performed. Then, given a new image, the deep learning model can predict the structural descriptor. Note that the input to the simulations is the structure, and the output of the simulations is the CBED patterns.
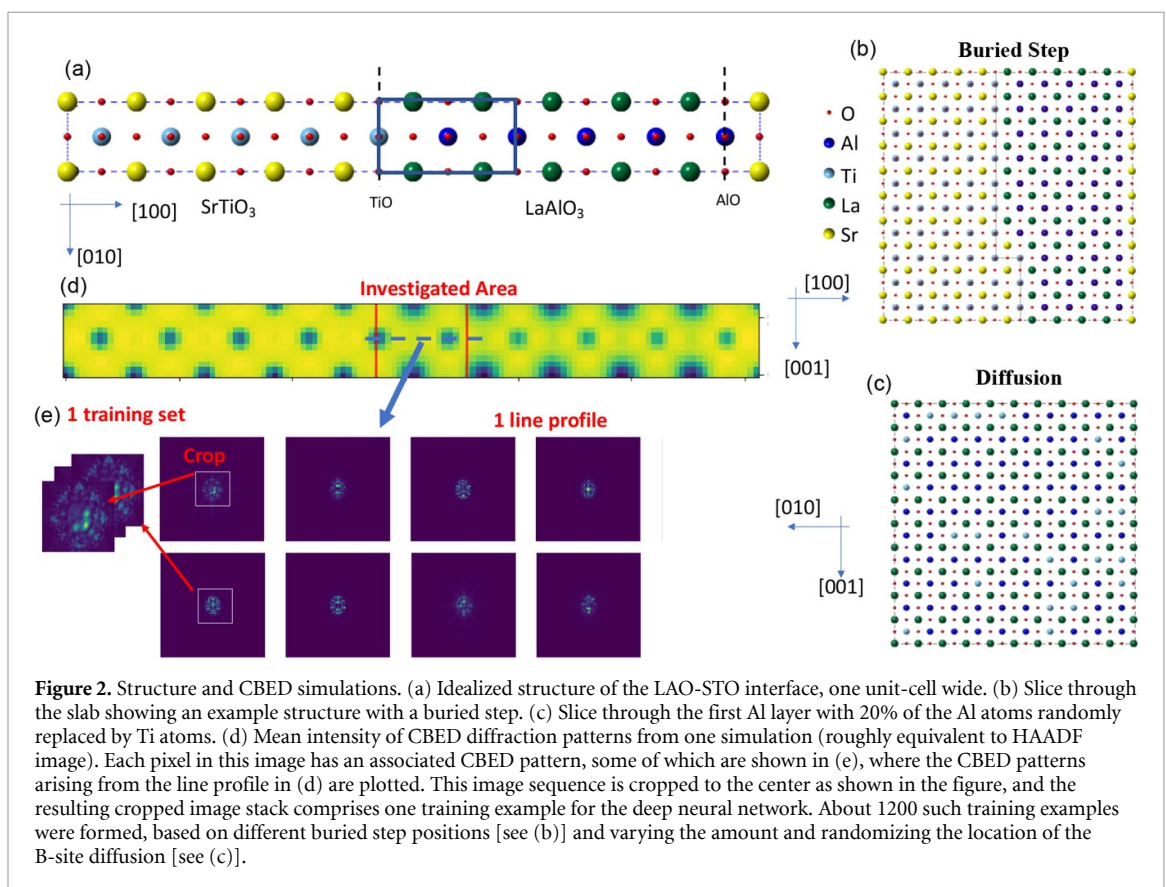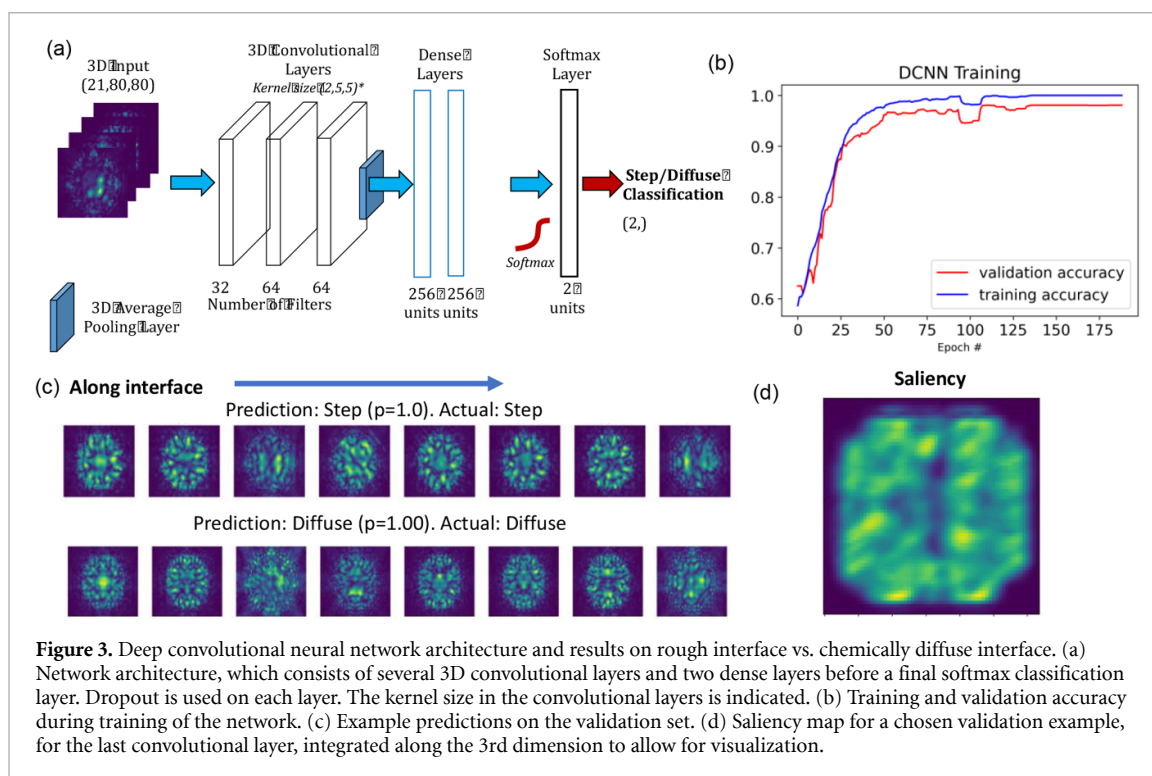

Figure 2. Structure and CBED simulations. (a) Idealized structure of the LAO-STO interface, one unit-cell wide. (b) Slice through the slab showing an example structure with a buried step. (c) Slice through the first Al layer with 20% of the Al atoms randomly replaced by Ti atoms. (d) Mean intensity of CBED diffraction patterns from one simulation (roughly equivalent to HAADF image). Each pixel in this image has an associated CBED pattern, some of which are shown in (e), where the CBED patterns arising from the line profile in (d) are plotted. This image sequence is cropped to the center as shown in the figure, and the resulting cropped image stack comprises one training example for the deep neural network. About 1200 such training examples were formed, based on different buried step positions [see (b)] and varying the amount and randomizing the location of the B-site diffusion [see (c)].

An example of the results of a simulation of a single structure is shown in figure 2(d), where we have computed the average intensity of each CBED pattern yield what would be roughly equivalent to a dark-field image. Since the dataset was large, and since the majority of the information is likely to reside at the interface, we restricted our study to the region shown by the bounded red lines in figure 2(d), which covers an area from the terminating TiO layer of the STO substrate to about one unit-cell into the LAO layer. To further

Figure 3. Deep convolutional neural network architecture and results on rough interface vs. chemically diffuse interface. (a) Network architecture, which consists of several 3D convolutional layers and two dense layers before a final softmax classification layer. Dropout is used on each layer. The kernel size in the convolutional layers is indicated. (b) Training and validation accuracy during training of the network. (c) Example predictions on the validation set. (d) Saliency map for a chosen validation example, for the last convolutional layer, integrated along the 3rd dimension to allow for visualization.

reduce the size to enable deep learning, we then cropped the CBED patterns to only the central portions, as these are expected to contain the majority of the information. Note that wavelet compression could also be used in place of this step for forming a reduced representation, but we found it unnecessary. The CBED patterns along the dotted blue line in figure 2(d) are shown in figure 2(e). After cropping of the image sequence, one arrives at a single stack of images associated with that particular structure, and more such sets were generated via changing the location of the line profile (i.e. moving the line across the interface).

As a first step, we attempted to determine whether a DCNN could take these sets of images and determine (classify) whether they originated from a rough interface (with a buried step), or from an interface that had some level of B-site diffusion. To do this, we used a DCNN with 3D convolutional layers followed by an average pooling layer, and then two dense layers, followed by a final softmax layer for the classification for the task. The full architecture is shown in figure 3(a). It is worth mentioning that the reason for the lack of pooling (and especially max pooling) after each layer, which is typical for most DCNNs is that this results information loss on relative positions of features with respect to each other [29]. This information is important in diffraction and indeed we found that including such layers dramatically reduced the network accuracy. We also note that we did not employ the regularly used architectures such as ResNet, Inception, VGG16, etc. The reason for this is twofold. First, we found that increasing the number of layers did not lead to increased (and in some cases actually decreased) performance, and secondly, we wanted to use the smallest possible network that would provide strong results. On the other hand, the lack of pooling in our architecture does increase memory requirements. We also utilized dropout on each layer (15%) for reducing overfitting and trained the network on two nodes of the Summit supercomputer using the Horovod distributed framework, for our model built in Keras [30] using the Tensorflow backend. Although we had ~7150 simulations of each class, deep learning typically requires larger datasets to be effective. We therefore utilized data augmentation, via addition of varying levels of Poisson noise as well as small rotations, as can be expected in real experiments to artificially inflate the data volume. The data was validated on a set of simulated data that was not shown to the model during the training phase, consisting of 1430 datasets selected at random and held back, and augmented during each epoch. We utilized standard stochastic gradient descent (SGD) (learning rate = 0.04, momentum = 0.0) optimizer and trained for 200 epochs, resulting in convergence of the model. SGD is an optimization algorithm that is standard practice in the deep learning community, and is an approximation to traditional gradient descent, well-suited to situations where it may be computationally too expensive to fully calculate the gradients [31].

The results of the DCNN as a function of training can be seen in figure 3(b). After about 110 epochs the accuracy is approaches ~97% on the validation set, suggesting that the network is almost perfectly capable of distinguishing between a rough interface from a chemically diffuse but sharp interface. An example of two predictions are shown in figure 3(c), where in the first case the DCNN predicts the image sequence shown

contains a step (with >99% certainty, as gauged from the softmax layer output), and in the second case, as a chemically diffuse interface (with 100% from the softmax layer output), both of which are correct. Here we note that even to a skilled expert in CBED pattern analysis, distinguishing between the two cases is not straightforward. In fact, the network was trained on patterns derived from structures of different thickness, which greatly affects the resulting diffraction due to additional scattering as the probe propagates through the crystal.
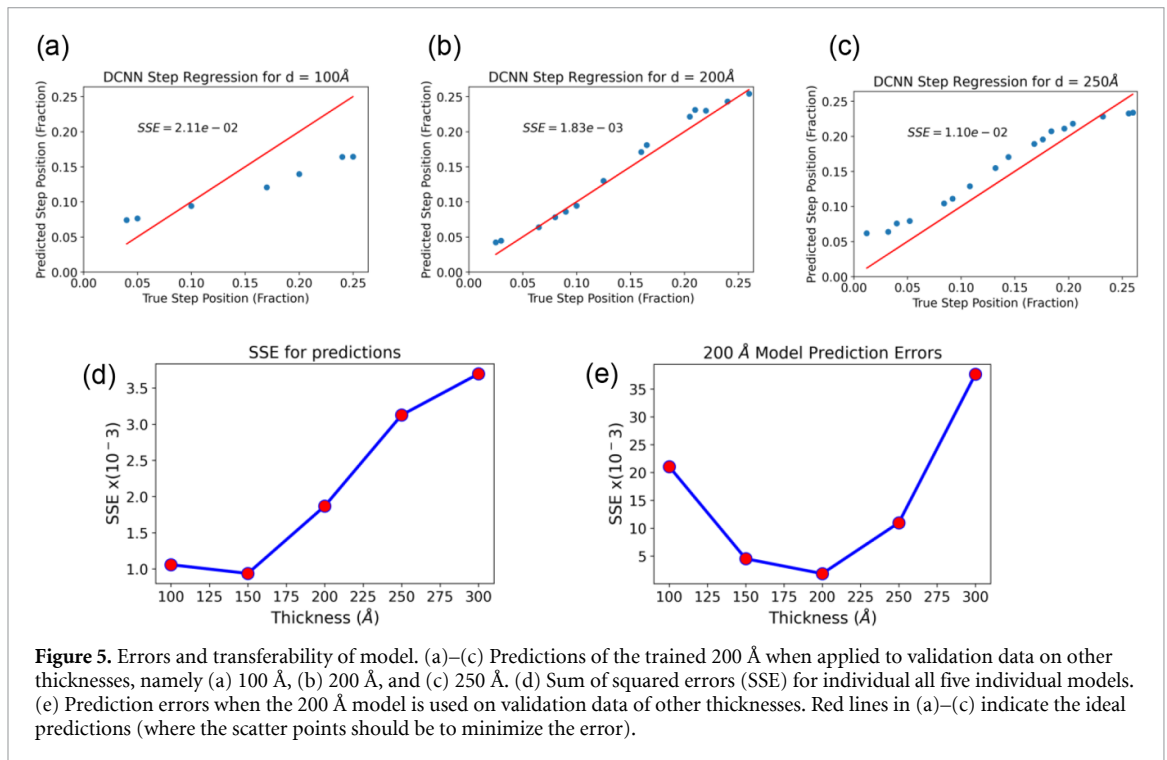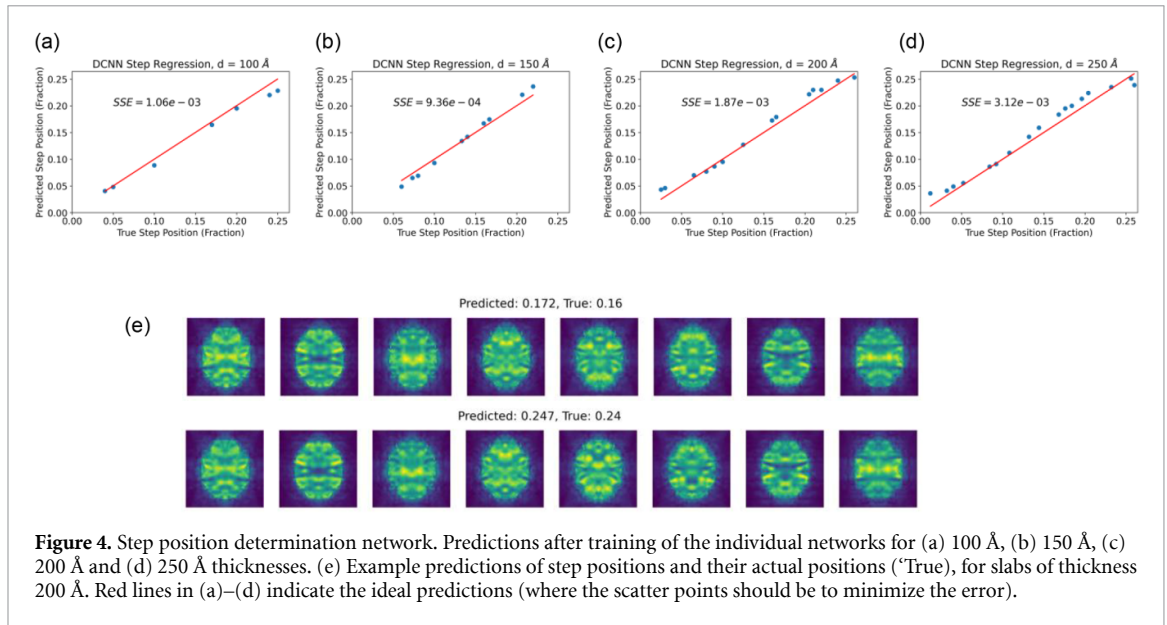
This raises an interesting question about how the network is coming to these conclusions. To gain insight, we explored the use of saliency maps [32] on the convolutional filters in the last convolutional layer of the network. The idea behind the saliency map is that it highlights regions of the image which cause the greatest change in the output, in this case, in the step/diffuse classification. The saliency map for a randomly selected validation image is shown in figure 3(d). Because this is a 3D image, we instead plot the integral along the 3rd dimension for ease of viewing. In fact, rather than focusing on a single part of the image, it seems many parts of the image appear important to the prediction.

Next, we explored whether we could utilize the same neural network architecture for a more challenging problem: for rough interfaces (those with buried steps), could the network be trained to determine where this buried interface resided? We used the same network architecture as in figure 3(a) with the exception that the last layer be changed to a linear activation as opposed to a softmax one, and the output being just one value (the step position, as a fraction of the specimen thickness). Two changes had to be made in order for the network training to converge: (1) the training set had to be combined such that the CBED patterns had to be averaged across the unit cell, unlike the individual line profile in figure 2(d), and (2) we had to train separate models for each thickness, unlike in the previous case where a single model could be trained for all thicknesses. Given the need to separate out data from different thicknesses, and the need to average over the unit cell, this resulted in substantial dataset reduction: for instance, we used in total 338 simulations for the training data and 208 simulations for the validation data for training the 200 Å model. The numbers for datasets reduce for the lower thickness models somewhat, due to less positions that the step can be in but increase for the models trained on the thicker specimens. As noted previously, the simulated data was augmented during training each epoch with addition of noise and small rotations, and the same was done to the validation data.

Shown in figures 4(a)–(d) are the results of the predicted vs. actual step position fractions for four different thicknesses (100 Å, 150 Å, 200 Å, 250 Å) for the trained models. The results show that for the most part, the regression models are effective at determining the step position for all the thickness values; however, this becomes more challenging at the higher thickness. To visualize the difficulty of this task, shown in figure 4(e) is the result of the network's predictions and the ground truth for two validation datasets for the 200 Å thickness. The network is correctly able to predict the step position, despite the fact that the differences are quite small and difficult to detect by eye. We note that the error in the predictions increases for increasing thickness as shown in figure 5(d). This agrees with our physical intuition, as the presence of multiple scattering in thicker crystals leads to more effective 'blurring', leading to less pronounced changes of the diffraction patterns with step location, and therefore making the step position determination more difficult. If the network was instead focused on some (unknown) artefact of the simulations, then it is hard to see why it would be more difficult to predict step positions for different thicknesses. Thus, we suggest that the network is indeed paying attention to the relevant features that distinguish between steps at different positions.

Next, we investigated the transferability of the approach to samples of other thickness. During any real experiment, there will be small thickness variations in the prepared specimens that would be unavoidable due to the nature of the milling process. We therefore tested the 200 Å model on simulated validation data for different thicknesses, with the results shown in figures 5(a)–(c). The plot of the model errors with thickness is shown in figure 5(e), and clearly shows that as the thickness deviates from 200 Å, the error increases, as would be expected. However, the error increases less for the thinner specimens than for the thicker ones, again agreeing with physical intuition. The error within 25 Å of the trained model is also reasonable; this suggests that small thickness variations (~5%–10% of the thickness simulated) can be tolerated in the experiment.

In summary, we have shown using simulated data that 4D STEM can be a powerful tool in distinguishing between different structures buried within cross-section samples, via leveraging of high-performance computing and DCNNs. We show that such neural networks have the ability to distinguish chemically diffuse interfaces in LAO-STO, from an interface that includes a buried step within the cross-sectional slice studied with very good reliability. In addition, we train a neural network with a similar architecture to estimate the location of the buried step, if it exists, and show that the estimate is accurate for simulated data in ~85% of cases. The method here is applied to 4D STEM but can be useful in any imaging situation where

Figure 4. Step position determination network. Predictions after training of the individual networks for (a) 100 Å, (b) 150 Å, (c) 200 Å and (d) 250 Å thicknesses. (e) Example predictions of step positions and their actual positions ('True'), for slabs of thickness 200 Å. Red lines in (a)–(d) indicate the ideal predictions (where the scatter points should be to minimize the error).

Figure 5. Errors and transferability of model. (a)–(c) Predictions of the trained 200 Å when applied to validation data on other thicknesses, namely (a) 100 Å, (b) 200 Å, and (c) 250 Å. (d) Sum of squared errors (SSE) for individual all five individual models. (e) Prediction errors when the 200 Å model is used on validation data of other thicknesses. Red lines in (a)–(c) indicate the ideal predictions (where the scatter points should be to minimize the error).

multiple structural possibilities exist but the inversion of the acquired images is complex, provided that large scale simulations of the different structural possibilities is feasible.

Acknowledgments

Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

Data availability statement

ORCID iD

R K Vasudevan ● https://orcid.org/0000-0003-4692-8579

References

[1] Ohtomo A and Hwang H Y 2004 A high-mobility electron gas at the LaAlO$_3$/SrTiO$_3$ heterointerface *Nature* **427** 423–6

[2] Perna P *et al* 2010 Conducting interfaces between band insulating oxides: the LaGaO(3)/SrTiO(3) heterostructure *Appl. Phys. Lett.* **97** 152111–3

[3] Brinkman A, Huijben M, Van Zalk M, Huijben J, Zeitler U, Maan J C, Van der Wiel W G, Rijnders G, Blank D H A and Hilgenkamp H 2007 Magnetic effects at the interface between non-magnetic oxides *Nat. Mater.* **6** 493–6

[4] Fechner M, Ostanin S and Mertig I 2009 Effect of oxidation of the ultrathin Fe electrode material on the strength of magnetoelectric coupling in composite multiferroics *Phys. Rev.* B **80** 094405

[5] Rondinelli J M, Stengel M and Spaldin N A 2008 Carrier-mediated magnetoelectricity in complex oxide heterostructures *Nat. Nanotech.* **3** 46–50

[6] Bousquet E, Dawber M, Stucki N, Lichtensteiger C, Hermet P, Gariglio S, Triscone J M and Ghosez P 2008 Improper ferroelectricity in perovskite oxide artificial superlattices *Nature* **452** 732-U4

[7] Garcia-Barriocanal J, Rivera-Calzada A, Varela M, Sefrioui Z, Iborra E, Leon C, Pennycook S J and Santamaria J 2008 Colossal ionic conductivity at interfaces of epitaxial ZrO$_2$: Y$_2$O$_3$/SrTiO$_3$ heterostructures *Science* **321** 676–80

[8] Borisevich A Y *et al* 2010 Suppression of octahedral tilts and associated changes in electronic properties at epitaxial oxide heterostructure interfaces *Phys. Rev. Lett.* **105** 087204

[9] He J, Borisevich A, Kalinin S V, Pennycook S J and Pantelides S T 2010 Control of octahedral tilts and magnetic properties of perovskite oxide heterostructures by substrate symmetry *Phys. Rev. Lett.* **105** 227203

[10] Jia C L, Mi S B, Faley M, Poppe U, Schubert J and Urban K 2009 Oxygen octahedron reconstruction in the SrTiO(3)/LaAlO(3) heterointerfaces investigated using aberration-corrected ultrahigh-resolution transmission electron microscopy *Phys. Rev.* B **79** 081405

[11] Gerra G, Tagantsev A K and Setter N 2007 Ferroelectricity in asymmetric metal-ferroelectric-metal heterostructures: a combined first-principles-phenomenological approach *Phys. Rev. Lett.* **98** 207601

[12] Kalinin S V, Borisevich A and Fong D 2012 Beyond condensed matter physics on the nanoscale: the role of ionic and electrochemical phenomena in the physical functionalities of oxide materials *Acs Nano* **6** 10423–37

[13] Kalinin S V and Spaldin N A 2013 Functional ion defects in transition metal oxides *Science* **341** 858–9

[14] Borisevich A Y *et al* 2012 Interface dipole between two metallic oxides caused by localized oxygen vacancies *Phys. Rev.* B **86** 140102

[15] Kim Y M, Morozovska A, Eliseev E, Oxley M P, Mishra R, Selbach S M, Grande T, Pantelides S T, Kalinin S V and Borisevich A Y 2014 Direct observation of ferroelectric field effect and vacancy-controlled screening at the BiFeO$_3$/LaxSr1-xMnO$_3$ interface *Nat. Mater.* **13** 1019–25

[16] Qiao L, Droubay T C, Kaspar T C, Sushko P V and Chambers S A 2011 Cation mixing, band offsets and electric fields at LaAlO$_3$/SrTiO$_3$(001) heterojunctions with variable La:Al atom ratio *Surf. Sci.* **605** 1381–7

[17] Cantoni C, Gazquez J, Miletto Granozio F, Oxley M P, Varela M, Lupini A R, Pennycook S J, Aruta C, Di Uccio U S and Perna P 2012 Electron transfer and ionic displacements at the origin of the 2D electron gas at the LAO/STO interface: direct measurements with atomic-column spatial resolution *Adv. Mater.* **24** 3952–7

[18] Kumar A, Ovchinnikov O, Guo S, Griggio F, Jesse S, Trolier-McKinstry S and Kalinin S V 2011 Spatially resolved mapping of disorder type and distribution in random systems using artificial neural network recognition *Phys. Rev.* B **84** 024203

[19] Ovchinnikov O S, Jesse S, Bintacchit P, Trolier-Mckinstry S and Kalinin S V 2009 Disorder identification in hysteresis data: recognition analysis of the random-bond-random-field ising model *Phys. Rev. Lett.* **103** 157203

[20] Cherukara M J, Nashed Y S and Harder R J 2018 Real-time coherent diffraction inversion using deep generative networks *Sci. Rep.* **8** 1–8

[21] Cherukara M J, Zhou T, Nashed Y, Enfedaque P, Hexemer A, Harder R J and Holt M V 2020 Real-time sparse-sampled ptychographic imaging through deep neural networks *arXiv preprint* (https://arxiv.org/abs/2004.08247)

[22] Laanait N, Romero J, Yin J, Young M T, Treichler S, Starchenko V, Borisevich A, Sergeev A and Matheson M 2019 Exascale deep learning for scientific inverse problems arXiv preprint (https://arxiv.org/abs/1909.11150)

[23] LeCun Y and Bengio Y 1995 Convolutional networks for images, speech, and time series *Handbook Brain Theory Neural Netw.* **3361** 1995

[24] Warusawithana M, Richter C, Mundy J, Roy P, Ludwig J, Paetel S, Heeg T, Pawlicki A, Kourkoutis L and Zheng M 2013 LaAlO3 stoichiometry is key to electron liquid formation at LaAlO$_3$/SrTiO$_3$ interfaces *Nat. Commun.* **4** 2351

[25] Zaid H, Berger M-H, Jalabert D, Walls M, Akrobetu R, Goble N, Gao X, Berger P, Fongkaew I and Lambrecht W 2018 Role of the different defects, their population and distribution in the LaAlO$_3$/SrTiO$_3$ heterostructure's behavior *J. Appl. Phys.* **123** 155304

[26] Allen L J and Findlay S 2015 Modelling the inelastic scattering of fast electrons *Ultramicroscopy* **151** 11–22

[27] Susner M A, Belianinov A, Borisevich A, He Q, Chyasnavichyus M, Demir H, Sholl D S, Ganesh P, Abernathy D L and McGuire M A 2015 High-T C layered ferrielectric crystals by coherent spinodal decomposition *ACS Nano* **9** 12365–73

[28] Yang H, Jones L, Ryll H, Simson M, Soltau H, Kondo Y, Sagawa R, Banba H, MacLaren I and Nellist P 2015 4D STEM: high efficiency phase contrast imaging using a fast pixelated detector *J. Phys.: Conf. Ser.* **644** 012032

[29] Vasudevan R K, Laanait N, Ferragut E M, Wang K, Geohegan D B, Xiao K, Ziatdinov M, Jesse S, Dyck O and Kalinin S V 2018 Mapping mesoscopic phase evolution during E-beam induced transformations via deep learning of atomically resolved images *Npj Comput. Mater.* **4** 30

[30] Chollet F 2015 Keras (https://github.com/fchollet/keras)

[31] Bottou L 2010 *Proceedings of COMPSTAT'2010* (Berlin: Springer) pp 177–86

[32] Simonyan K, Vedaldi A and Zisserman A 2013 Deep inside convolutional networks: visualising image classification models and saliency maps *arXiv preprint* (https://arxiv.org/abs/1312.6034)