3022
doi:10.1017/S1431927622011308

Microsc. Microanal. 28 (Suppl 1), 2022

Versatile Automated Domain Mapping of 4D-STEM Data Utilizing ML Algorithms and Bayesian Statistics

A. Bridger[1,2,3], M. Danaie[1], K. Butler[2], T. Wood[2] and W. I. F. David[2,3]

[1.] Diamond Light Source, Didcot, Oxfordshire, United Kingdom
[2.] Science and Technology Facilities Council, Swindon, Wiltshire, United Kingdom
[3.] University of Oxford, Oxfordshire, United Kingdom

This work outlines the development of a workflow that combines both informed and data-driven approaches to process and classify 4D-STEM data utilising signal dimensionality reduction obtained from unsupervised machine learning algorithms. Using these 'latent space' representations of the scanned region, we can then apply clustering algorithms to identify regions of similar character. By combining the results of multiple autonomous computations, we would be able to map domains within the crystal and associate a level of confidence to given classifications. This confidence is then used to iteratively improve the learning algorithms and the performance of their dimensionality reduction. Incorporating this as a pre-processing unsupervised workflow would drastically improve the ability to characterise the nano-scale structure of materials, both through production of significantly signal-boosted diffraction data and reduction in laborious manual investigation.

With the advent of fast direct-electron detectors, capturing large scan arrays of electron diffraction data is becoming more and more accessible. The Scanning Electron Nanobeam Diffraction (SEND) technique lends itself well to microscope automation and can cover areas in the micrometres length scale and hence, while still capturing microstructural details not accessible via powder x-ray diffraction (PXRD), provides statistically significant insight into the crystallographic defects and domain morphology in the microstructure. Larger probe sizes and lower magnifications typically used in SEND also result in reduced electron dose on the sample, compared to atomic-resolution STEM, making this technique suitable for beam-sensitive phases. The quantity of SEND data collected during a day on the microscope can run well above 1TB. This is far more information than can be comprehensively processed by hand within a reasonable time frame.

As a case study, we are implementing this approach to study the complex microstructure that drives leading Sodium-ion battery materials. Our materials are based on the same layered metal oxide structures as the commercial lithium NMC and NCA cathodes but have a more complex structural polymorphism. A comprehensive knowledge of the different structures is essential for our understanding of these materials and, due to the potential complexity of phase disposition, these materials are perfect case studies for our unsupervised domain mapping approach.

For the SEND data collection, we aligned the JEOL Grand-ARM microscope at 300 kV with the probe corrector optics turned off, allowing us to reduce the probe convergence semi-angle to around 1 mrad. We collected the diffraction data on a MerlinEM quad detector (Quantum Detectors) with each dataset reshaping to 256x256 in the probe scanning plane and 515x515 on the diffraction plane.

Our proposed workflow is:

1. Train a Variational AutoEncoder (VAE) to embed the patterns in a 2D latent space
2. Distort the latent space to improve class separation and aid clustering

3. Cluster the embedded points
4. Repeat for multiple distortions and identify the patterns which are inconsistently classified within the current latent representation
5. Adjust the VAE training process to improve the embedding of these regions
6. Continue iterating until the model produces a domain map with the desired level of consistency

The primary method for dimensionality reduction is using a Variational Auto Encoder (VAE). The VAE offers two key advantages which are leveraged in our workflow: versatility and continuity. The autoencoder can be trained on any 4D-STEM data and is not beholden to any single lattice type, allowing this to be applied to a variety of diverse systems. The autoencoder is also advantageous over other popular dimensionality reduction techniques such as t-distributed Stochastic Neighbour Embedding, Multidimensional Scaling or Principal Component Analysis, due to the continuous nature of the latent representation it generates. This allows us to leverage insight from the regions of the latent space that don't contain embedded sample points and is an incredibly useful property necessary for applying transforms to the latent space to improve class separation. This class separation technique allows us to cluster the low dimensionality representations more effectively than any of the other approaches we have tested.

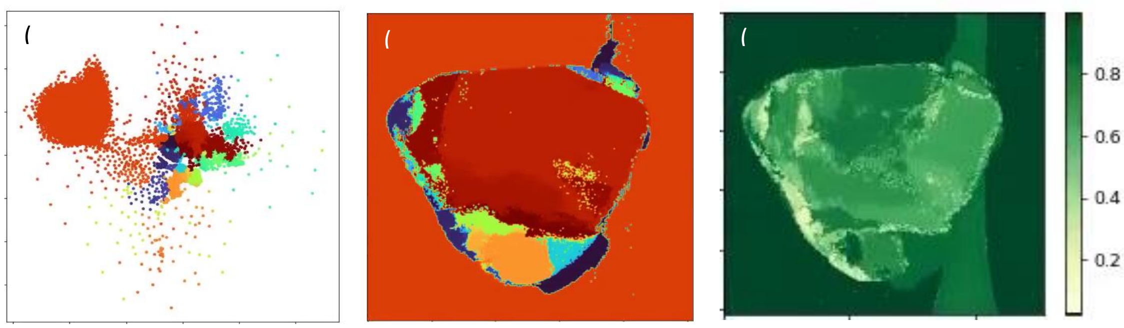

Figure 1. (a) the clustered two-dimensional representation of the diffraction data, (b) the clustering applied to the probe positions in the navigation plane, (c) the intermediate confidence map used to identify regions that were being inconsistently classified.

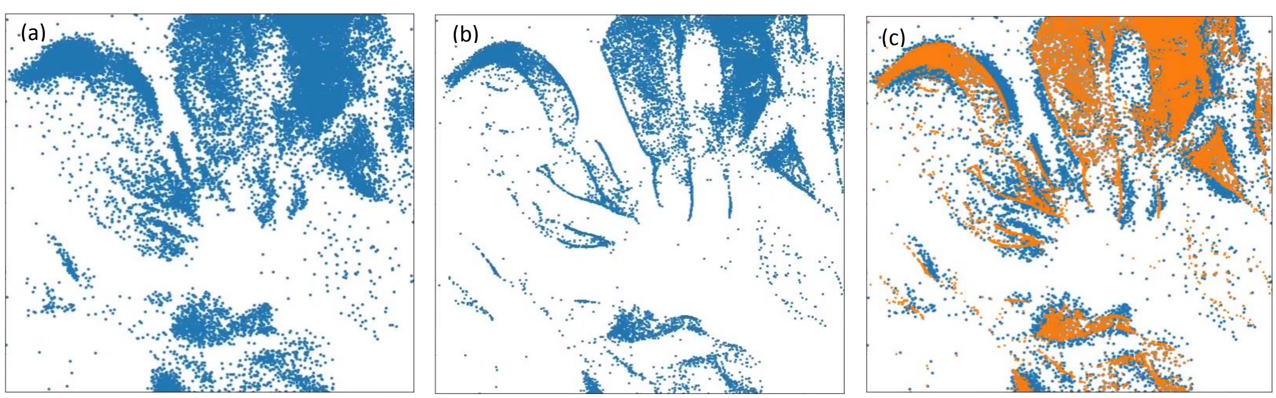

Figure 2. (a) example sub-region of latent space encoding showing the diffraction encoding position, (b) an example latent space transformation showing improved class separation, (c) a comparison of the two distributions.

https://doi.org/10.1017/S1431927622011308 Published online by Cambridge University Press