doi:10.1017/S1431927622011242
Microsc. Microanal. 28 (Suppl 1), 2022

Exploring Motifs and Their Hierarchies in Crystals via Unsupervised Learning

Jiadong Dan[1,2], Xiaoxu Zhao[3], Qian He[4], N. Duane Loh[1,2,5*] and Stephen J. Pennycook[6,7*]

[1.] NUS Centre for Bioimaging Sciences, National University of Singapore, Singapore.
[2.] Dept. of Biological Science, National University of Singapore, Singapore.
[3.] School of Materials Science and Engineering, Nanyang Technological University, Singapore.
[4.] Dept. of Materials Science and Engineering, National University of Singapore, Singapore.
[5.] Dept. of Physics, National University of Singapore, Singapore.
[6.] Dept. of Materials Science and Engineering, University of Tennessee, Knoxville, TN, USA.
[7.] School of Physical Sciences, University of Chinese Academy of Sciences, Beijing, China.
* Corresponding author: duaneloh@nus.edu.sg (N.D.L); stephen.pennycook@cantab.net (S.J.P).

The confluence of materials sciences and artificial intelligence (AI) has been heralding a surge of research on functional materials discovery and property predictions. Machine learning, as one of the most essential techniques in AI, has proved its superhuman ability in guiding chemical synthesis [1], enhancing computational chemistry [2], targeting the discovery of new catalysts [3] and more. Despite its success in the field of materials discovery, complementary research in machine learning for materials characterization is still limited and primarily focuses on supervised deep learning models or direct adaptation of well-developed models to experimental datasets. Here, we present a machine learning framework that rapidly extracts and reports an interpretable hierarchy of complex structural motifs from atomically-resolved images [4].

Figure 1 shows one of the key developments in our framework: Zernike Polynomials (ZPs) are demonstrably more efficient than principal component analysis (PCA) in improving clustering performance in a low dose rate. We used an ADF-STEM image of monolayer $MoSe_2$ (Fig. 1a) and its lower-dose versions to explore the classification performance of ZPs and PCA. Both Fowlkes–Mallows index (Fig. 1b) and adjusted mutual information (Fig. 1c) show that ZP-based classification can be extended to dose rates that are 1/50 of the original image (effective dwell time 200 ns). Fig. 1d shows the average reconstructed motifs from the lower-dose image (1/50). This means we can extend the linear dimensions of our 2D scanning area by at least 7 times during the same acquisition cycle without sacrificing the accuracy of structure analysis. Such an increase will give us higher statistical power in understanding complex, disordered samples.

Deeply rooted in our framework is the critical notion of structural motifs. Motifs identified by our framework can *coarse-grain* and hence simplify the staggeringly large space of possible atomic configurations. While this coarse-graining might seem at odds with the push to increase resolution, many key insights about a disordered sample (*e.g.*, defects, inclusions, frustration) are already evident without sub-Angstrom resolution. Such coarse-grained motifs are useful for further classifying the higher resolution differences within each motif class owing to local strain variations.

Figure 2 illustrates how motif hierarchy in a polyoxometalate Mo-V-Te-Nb-oxide sample reveals novel but frustrated phases. Fig. 2a shows an ADF-STEM image of Mo-V-Te-Nb-oxide which we identified as a type-4 monohedral pentagonal structure consisting of distorted empty/filled pentagonal units. These basic motifs can be used to construct higher-level complex larger motifs as shown in Fig. 2b. The motif hierarchy presented reveals the frequent larger repeating units, which suggests fresh clues about

preferential attachment and free-energy barriers during multi-step nucleation and growth of ordered structures. The discovered persistent motifs can even be used to create initial models for downstream modeling.

To conclude, we have described a framework that exploits the spatial context between simple motifs to learn a hierarchical composition of higher-level motifs. The capability to algorithmically construct structural hierarchy from microscopy images without human supervision has the potential to alter how materials scientists characterize and interpret complex materials structures, usually not readily evident to human eyes, in areas like oxide thin films, metal halide perovskite, colloidal crystal [4].

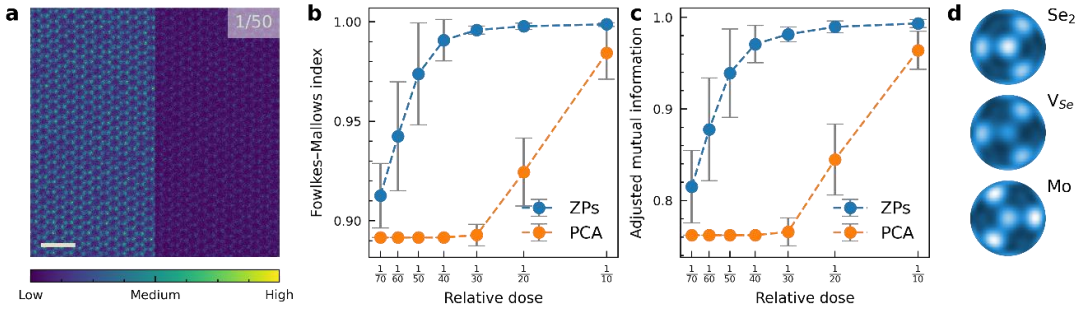

Figure 1. Zernike representations improve clustering separation in a low-dose setting. (a) ADF-STEM image of a monolayer $MoSe_2$ (scale bar: 1nm), whose right half is replaced by a low dose version (1/50 of the left half). (b) Fowlkes–Mallows index vs. relative dose using ZPs and PCA. (c) Adjusted mutual information vs. relative dose using ZPS and PCA. (d) Average reconstructed patches from the lower dose image (1/50).

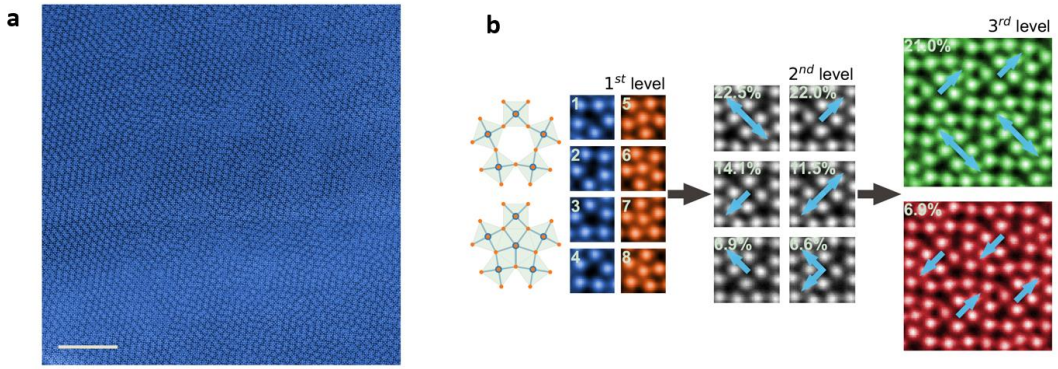

Figure 2. Hierarchy of motifs in polyoxometalate Mo-V-Te-Nb-oxide. (a) ADF STEM image of the complex metal oxide with no apparent long-range order. (scale bar: 10nm) (b) Hundreds of larger motifs can be composed of these filled/hollow pentagonal motifs in a 3-level hierarchy. Turquoise arrows indicate the presence of filled pentagonal units.

References:

[1] JM Granda et al., Nature **559** (2018), doi:10.1038/s41586-018-0307-8
[2] F Brockherde et al., Nature Communications **8** (2017), doi:10.1038/s41467-017-00839-3
[3] AF Zahrt et al., Science **363** (2019), doi:10.1126/science.aau5631
[4] S. J. P. acknowledges funding from Singapore Ministry of Education Tier 1 grant R-284-000-172-114, Tier 2 grant R-284-000-175-112. N.D.L acknowledges funding support from the National Research Foundation (grant number NRF-CRP16-2015-05), and the NUS Early Career award (A-0004744-00-00).

https://doi.org/10.1017/S1431927622011242 Published online by Cambridge University Press