Exploring Motifs and Their Hierarchies in Crystals via Unsupervised Learning

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

1. NUS Centre for Bioimaging Sciences, National University of Singapore, Singapore. 2. Dept. of Biological Science, National University of Singapore, Singapore. 3. Dept. of Materials Science and Engineering, National University of Singapore, Singapore. 4. Dept. of Materials Science and Engineering, University of Tennessee, Knoxville, TN, USA. 5. School of Physical Sciences, University of Chinese Academy of Sciences, Beijing, China.

Introduction

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].

Key Challenges

1. I don't understand what my ML model is learning! How to represent structural motifs with human-interpretable features?

2. Unsupervised identification of motifs: How to improve cluster separability?

3. Can we generalize to different types of samples and modalities (e.g., AFM, STM, STEM, *etc.*)? How to keep the framework flexible?

4. How structural motifs are randomly placed? How do they self-organize into hierarchies?

Structural motif representations

Challenge 1

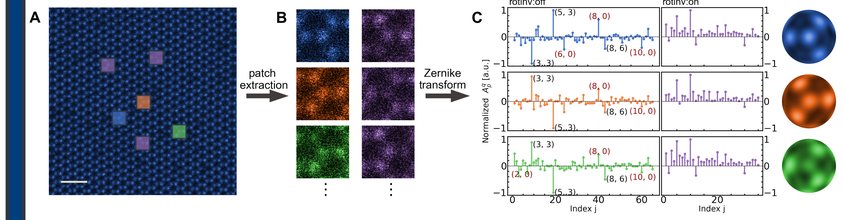


Fig. 1. Interpretable and flexible atomic structural motifs representation via Zernike polynomials (ZPs). **(A)** ADF-STEM image of monolayer $MoSe_2$ (scale bar: 0.5 nm). **(B)** Automatically extracting atom-centered image patches from panel (A). **(C)** Computing the Zernike features within the image patches in panel (B) either in their rotationally invariant representation (rotinv:on) or retaining rotation information (rotinv:off).

Two-Stage Force-relaxed clustering

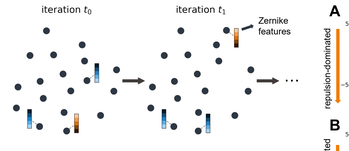
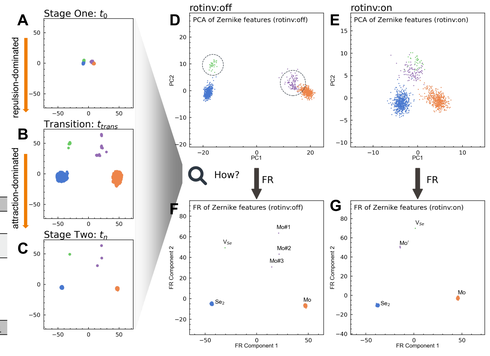


Fig. 2. Illustration of force-relaxed algorithm.

	Attraction force	Repulsion force	Remarks
LargeVis	$\frac{2}{(1+d^2)}$	$\frac{2\gamma}{(1+d^2)(d^2+\epsilon)}$	$\gamma = 7, \epsilon = 0.1$
UMAP	$\frac{2ab\, d^{2b-2}}{(1+a\, d^{2b})}$	$\frac{2\gamma b}{(1+a\, d^{2b})(d^2+\epsilon)}$	$\gamma = 1, \epsilon = 0.001$ $a = 1.58, b = 0.90$
Two-stage relaxed clustering (this work)	$\frac{\alpha}{1+d^n}$	$\frac{\beta}{1+d^m}$	$m \geq 0, n \geq 0$ $\alpha > 0, \beta > 0$

Here d refers to the distance between two PCA-reduced features: $\boldsymbol{d} = ||\boldsymbol{Y}_u - \boldsymbol{Y}_v||$.

Table 1. A summary of force functions in different iterative clustering algorithms



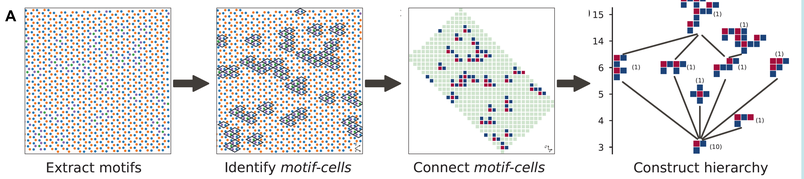
Challenge 2

Fig. 3. PCA and Force-relaxed (FR) layouts of Zernike features from monolayer $MoSe_2$ sample. From **(A)** to **(B)**, repulsion-dominated stage. From **(B)** to **(C)**, attraction-dominated stage. **(D)** and **(E)** project the patches' Zernike representations in their first two principal components. **(F)** Applying a force-relaxed clustering algorithm acting on these Zernike features automatically classifies them into different atomic structural motifs; (**G**) clustering on the rotationally invariant representations (rotinv:on) neatly groups features centered on Mo atoms together.

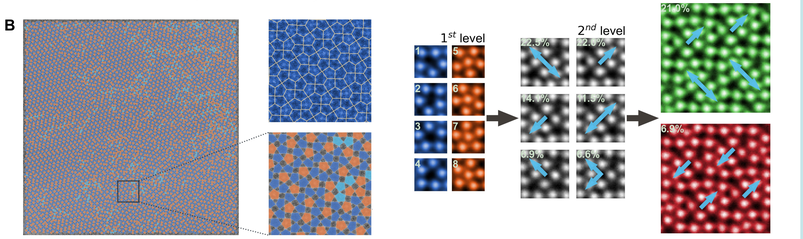
Combination of ZPs and FR gives more flexibility (Challenge 3)

Motif hierarchy in sample complexity spectrum

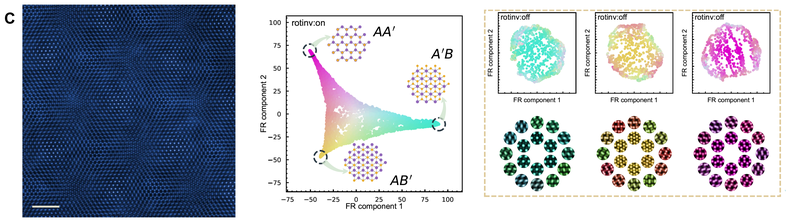
Simple motifs, simple hierarchy



Simple motifs, complex hierarchy



Complex motifs, "continuous hierarchy" becomes manifold



Increasing Sample Complexity

Fig. 4. Hierarchies of motifs in samples of increasing structural complexity. **(A)** Construction of hierarchy of *motif-cells* found in the monolayer $MoSe_2$ sample. **(B)** Hierarchy of motifs in polyoxometalate Mo-V-Te-Nb-oxide. **(C)** Structural motifs in twisted bilayer MoS_2 form spatial hierarchy reflected by a continuous manifold.

Challenge 4

Conclusions

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.

Acknowledgement

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).

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. J Dan et al., *Science Advances* **8** (2022), doi: 10.1126/sciadv.abk1005

Online Materials →

SCAN ME