Bridging Contextual Information in Deep Learning for Structural Defect Classification

Jiadong Dan [1] **Cheng Zhang** [2] **Leyi Loh** [3] **Goki Eda** [1] **N. Duane Loh** [1]

[1]*National University of Singapore, Singapore, 119077* [2]*City University of Hong Kong, Hong Kong SAR* [3]*University of Cambridge, United Kingdom*. Correspondence to: Jiadong Dan dan.jiadong@nus.edu.sg, Duane N. Loh duaneloh@nus.edu.sg.

1. Introduction

Accurate and rapid identification of structural defects in materials using atomically resolved scanning transmission electron microscopy (STEM) is essential for understanding structure-property relationships. While deep learning models have been applied to classify atomic defects from atomic-resolution STEM images [1, 2], they often struggle to generalize across different materials. Specifically, deep neural networks (DNNs) trained on one material frequently underperform when applied to unseen structures from other materials due to out-of-distribution challenges, limiting their practical utility.

What limits the generalizability of current models? A major factor is their failure to explicitly incorporate essential contextual information—such as chemical composition and experimental conditions—into the training process. Although this information is often available, models relying solely on pixel-based inputs are prone to misclassification, as images from distinct materials can appear visually similar despite having fundamentally different chemical compositions. For instance, simulated [3] high-angle annular dark field (HAADF) images of monolayer WTe_2 and $NbSe_2$ at 80 keV exhibit nearly indistinguishable contrast, as shown in Figure 1. This highlights the ill-posed nature of classification based purely on pixel intensities. Directly integrating contextual information is therefore crucial for achieving accurate and reliable defect identification across diverse material systems.

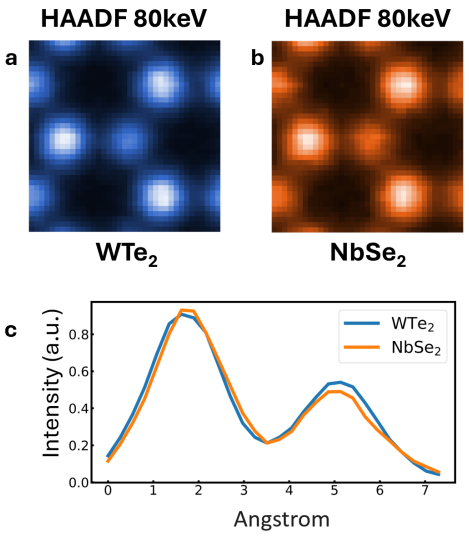

Fig. 1: **Image contrast alone is insufficient to reliably distinguish atomic defects in materials. a** and **b**, Simulated HAADF images of monolayer WTe_2 and $NbSe_2$ using an electron beam at 80 keV. **c**, The images in panels a and b exhibit indistinguishable contrast.

Methodology

To overcome these limitations, we integrate contextual information directly into our classification process, as shown in Figure 2a. Each image patch is associated with chemical composition, electron beam energy, and detector type that will be encoded into a feature vector via a context encoder. In this process, chemical composition is transformed into normalized electron shell configuration counts, while electron beam energy and detector type are represented as one-hot vectors. This engineered representation captures essential non-visual data, enabling the model to consider both chemical composition and imaging conditions that affect defect identification.

In parallel, we transform the image patches using a contrast encoder that converts them into a set of Zernike vectors [4]. Here, each patch is approximated as a linear combination of Zernike polynomials, with the resulting coefficients forming a vector representation. These coefficients capture rotationally invariant features, effectively summarizing pixel intensity patterns while preserving critical structural details. This approach ensures that the encoder extracts robust geometric and contrast-based features from atomic-resolution images, complementing the contextual information for improved classification performance.

The outputs from both the *context* and *contrast* encoders are then concatenated to form a combined representation. This combined feature set integrates both visual and contextual information, allowing the classifier to identify atomic defect types across a wide range of materials.

Results and Discussion

We demonstrate the effectiveness of our proposed method by training an attention model [5] on simulated images [3] of 96 distinct cases of 1H MX_2 monolayer transition metal dichalcogenides (TMDs) with various dopants, where M denotes the transition metal and X the chalcogen. The model achieves 96% accuracy on both training and testing datasets. Figure 2b shows the accuracy curves throughout the training process. Notably, the test accuracy marginally surpasses training accuracy as a result of

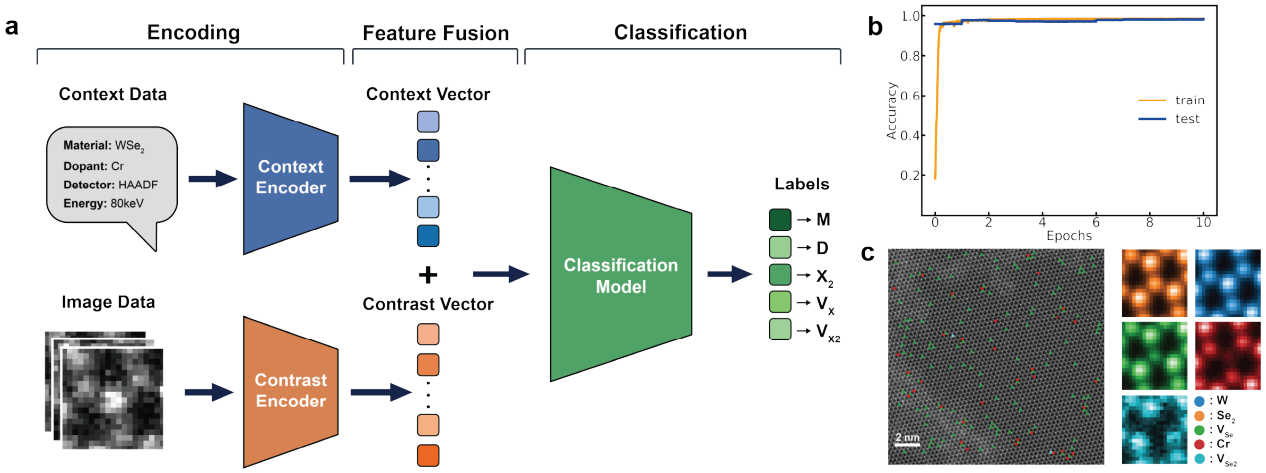

Fig. 2: **The architecture of the multimodal learning framework and classification results on experimental data. a**, The context encoder processes experimental context data, while the contrast encoder processes image data, converting both into one-dimensional feature vectors. These vectors are concatenated and input into the classification model, which categorizes them into one of five atomic column types: metal, dopant, X_2, single vacancy, or double vacancy. **b**, Accuracy curves of the defect attention model throughout the training process. **c**, Classification result of monolayer WSe_2 doped with Cr.

the stochastic regularization introduced by dropout during training [6].

We also developed an interactive graphical user interface (GUI) to streamline data labeling from experimental images. To establish trusted labels, we implemented a consensus-based voting system that aggregated annotations from domain experts. With these verified labels, our trained model was then applied to high-angle annular dark-field (HAADF) images of monolayer WSe_2 doped with Cr, V, Mn, and Co, yielding classification outcomes that align with expert assessments. A representative example is shown in Figure 2c. Notably, although our model was trained exclusively on simulated data, it generalizes effectively to experimental images.

Our results show that integrating domain-specific context enables a single model to generalize across diverse materials, overcoming the out-of-distribution challenges that limit conventional deep neural networks. This work advances deep learning for microscopy and microanalysis, shifting from task-specific models toward a unified framework for defect classification across varied material systems.

Acknowledgments

The authors acknowledge funding support from the National Research Foundation (grant number NRF-CRP16-2015-05), and the NUS Early Career award (A-0004744-00-00). This project is supported by the Eric and Wendy Schmidt AI in Science Postdoctoral Fellowship, a Schmidt Sciences program.

References

[1] Kangshu Li, Xiaocang Han, Yuan Meng, Junxian Li, Yanhui Hong, Xiang Chen, Jing-Yang You, Lin Yao, Wenchao Hu, Zhiyi Xia, Guolin Ke, Linfeng Zhang, Jin Zhang, and Xiaoxu Zhao. Single-image-based deep learning for precise atomic defect identification. *Nano Lett.*, 24(33):10275–10283, 21 August 2024.

[2] Maxim Ziatdinov, Ayana Ghosh, Chun Yin (tommy) Wong, and Sergei V Kalinin. AtomAI framework for deep learning analysis of image and spectroscopy data in electron and scanning probe microscopy. *Nature Machine Intelligence*, 4(12):1101–1112, 8 December 2022.

[3] Jacob Madsen and Toma Susi. The abTEM code: transmission electron microscopy from first principles. *Open Res Europe*, 1:24, 24 March 2021.

[4] Jiadong Dan, Xiaoxu Zhao, Shoucong Ning, Jiong Lu, Kian Ping Loh, Qian He, N Duane Loh, and Stephen J Pennycook. Learning motifs and their hierarchies in atomic resolution microscopy. *Sci Adv*, 8(15):eabk1005, 15 April 2022.

[5] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł Ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I Guyon, U V Luxburg, S Bengio, H Wallach, R Fergus, S Vishwanathan, and R Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 5998–6008. Curran Associates, Inc., 2017.

[6] Nitish Srivastava, Geoffrey E Hinton, A Krizhevsky, I Sutskever, and R Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.*, 15(1):1929–1958, 2014.