Bridging Supervised Segmentation and Unsupervised Anomaly Detection for Automated Defect Analysis in 2D Transition Metal Dichalcogenides

Cobi Allen[1], Mostafa A. Mostafa[2], Nannan Zhang[3], Navid Haghmoradi[4]
[1]University of Cambridge, United Kingdom
[2]Escola Politècnica Superior, University of Girona, Girona, Spain
[3]University of Exeter, Exeter, United Kingdom
[4]Karlsruhe Institute of Technology, Karlsruhe, Germany

Abstract

Two-dimensional transition metal dichalcogenides (TMDs) are promising candidates for next-generation quantum materials due to their multi-element composition and tunable structures. In this work, we focus on quaternary monolayer TMD alloys composed of Mo, W, S, and Se, using high-angle annular dark-field (HAADF) scanning transmission electron microscopy (STEM) to generate atomic-resolution image datasets. We implemented supervised and unsupervised machine learning techniques to automatically map atomic structures and detect defects in these images, creating a scalable, data-driven framework for analyzing 2D TMD materials. This approach provides predictive guidance for the rational design of functional quaternary alloys for spintronic, electronic, and optoelectronic applications.

Background

Transition metal dichalcogenides (TMDs) are extensively investigated two-dimensional materials with tunable electronic properties [1,4]. Beyond binary and ternary compositions, multi-element TMD alloys exhibit hybrid characteristics analogous to high-entropy systems, offering enhanced lattice stability and improved defect tolerance [2].

In TMD systems, local defects, atomic clusters, and nanostructures strongly influence material properties [3]. Atomic-resolution HAADF-STEM and image stacks [5] provide a crucial experimental tool for investigating atomic-scale structural dynamics. However, traditional experiments struggle to capture large-scale, high-resolution, and time-resolved information, and conventional analysis cannot efficiently identify defect patterns from atomic-scale STEM datasets, limiting our research on alloying mechanisms and defect evolution in multi-element TMDs.

In this project, we aim to automate the identification and characterisation of atoms and defects in HAADF-STEM images using supervised and unsupervised machine learning approaches.

Supervised Analysis for Defects classification in 2D material STEM datasets

To enable robust atomic characterization, we developed a hybrid supervised framework combining semantic segmentation with physics-informed classification. First, large-scale HAADF-STEM images were tiled into 512×512 patches, where a U-Net convolutional neural network (Figure 1) was trained to generate dense semantic masks of atomic columns, achieving an F1-score of 0.94 for atom detection. While effective for localization, segmentation alone lacked chemical specificity for isoelectronic species. To resolve this, we

implemented a second-stage Physics-Informed Classifier using the AtomAI framework [6]. Individual atomic patches were extracted and passed through a custom constraint layer that enforced crystallographic rules—restricting metal species to A-sites and chalcogens to B-sites based on graph topology. This "physics lockdown" eliminated chemically impossible predictions, boosting classification accuracy from ~65% to 94.2% and enabling the automated discovery of specific point defects (V_{Mo} and V_S) that standard segmentation could not distinguish.

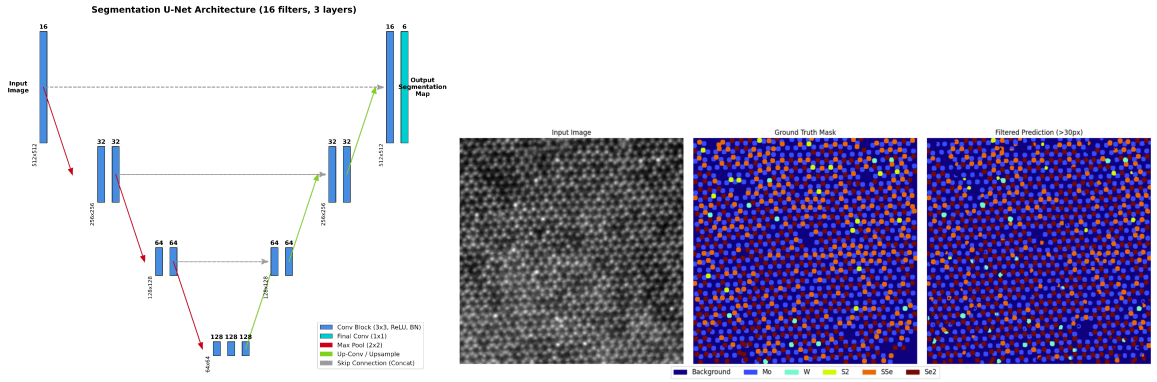

Figure 1. *Left:* Schematic of the modified U-Net architecture used for 6-class atomic segmentation.*Right:* Model predictions for a single HAADF-STEM image.

Unsupervised Analysis of Atomic-Scale Defect Dynamics

To complement supervised analysis, we apply an unsupervised, data-driven approach to image stacks to explore atomic-scale dynamics without predefined labels. Atomic columns are tracked across sequential images, and atom-centered patches are analyzed using a self-supervised convolutional autoencoder to learn compact representations of local atomic environments. Feature tracking over time defines an atomic instability metric that identifies dynamically changing sites. Spatial density analysis shows that these atoms are not randomly distributed but form localized, defect-rich regions within the lattice (Figure X). This unsupervised workflow provides complementary insight into atomic-scale defect dynamics and helps identify areas of structural heterogeneity.

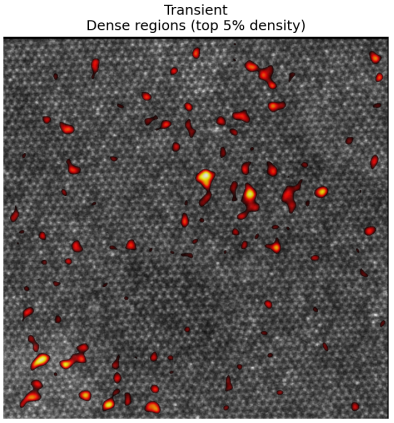

Figure 2. Spatial demonstration of dense regions formed by dynamically changing atomic sites environment identified using unsupervised analysis.

References

[1] Alboteanu, G., Mordehai, D., Ya'akobovitz, A. Mechanical modulation of 2D transition metal dichalcogenide alloys. Nanoscale, 2025, 17(11), 6512-6521.

[2] Schweidler, S., Botros, M., Strauss, F., Wang, Q., Ma, Y., Velasco, L., Marques, G., Sarkar, A., Kübel, C., Hahn, H., Aghassi-Hagmann, J., Brezesinski T., Breitung, B. High-entropy materials for energy and electronic applications. Nature Reviews Materials 9.4, 2024, 266-281.

[3] Haider, G., Sampathkumar, K., Verhagen, T., Nádvorník, L., Sonia, F. J., Valeš, V., Sýkora, J., Kapusta, P., Němec, P., Hof, M., Frank, O., Chen, Y., Vejpravová, J., Kalbáč, M. Superradiant emission from coherent excitons in van der Waals heterostructures. Advanced Functional Materials, 2021, 31(29), 2102196.

[4] Hanedar, B. A., Onbaşlı, M. C. Defect dependent electronic properties of two-dimensional transition metal dichalcogenides (2H, 1T, and 1T′ phases). Physical Chemistry Chemical Physics, 2025, 27(4), 1809-1818.

[5] Bampoulis, P., Sotthewes, K., Siekman, M. H., Zandvliet, H. J. Local Conduction in Mo x W1–x Se2: The Role of Stacking Faults, Defects, and Alloying. ACS applied materials & interfaces, 2018, 10(15), 13218-13225.

[6] Ziatdinov, M., Ghosh, A., Wong, C. Y., & Kalinin, S. V. (2022). AtomAI framework for deep learning analysis of image and spectroscopy data in electron and scanning probe microscopy. *Nature Machine Intelligence*, *4*(12), 1101–1112.