# Automated Detection of Regions of Interest in Atomic-Resolution STEM Images Using Patch-Based VGG16

R. A. W. Ayyubi[1], Shoaib Masood[1], Fatemeh Karimi[1], Zahira El Khalidi[1], Jaeyeon Jo[1]
[1] University of Illinois Chicago, Chicago, IL, USA

## Abstract

Scanning transmission electron microscopy (STEM) is a powerful tool for probing atomic-scale defects and local variations that govern material functionalities; however, identifying regions of interest (ROIs) remains a major bottleneck for its efficiency requiring extensive expert effort. We propose an automated ROI detection approach using patch-based classification with a VGG16 convolutional neural network, in which large field-of-view atomic-resolution images of CdTe and $SrTiO_3$ are divided into small patches and classified by type. Pre-trained on natural images, VGG16 successfully extracts meaningful features from patches containing approximately 2–4 unit cells, despite limited training datasets, enabling effective ROI determination in both supervised and unsupervised settings. This approach represents a step toward the automation of microscopy while reducing reliance on expert judgment.

## Purpose

The purpose of this work is to advance the automation and scalability of atomic-resolution STEM by reducing reliance on manual, expert-driven identification of regions of interest. While modern STEM instruments can rapidly acquire large volumes of high-resolution data, interpretation remains a key bottleneck that limits throughput, reproducibility, and real-time experimental decision-making. By establishing a generalizable, patch-based framework built on VGG-16–based feature learning [1], this study demonstrates detailed defect classification in CdTe STEM images and extends the same methodology to $SrTiO_3$, illustrating its applicability across distinct material systems. Together, these results provide a foundation for data-driven, adaptive microscopy workflows and support a transition from passive image acquisition toward closed-loop and autonomous atomic-scale characterization.

## Results & Discussions

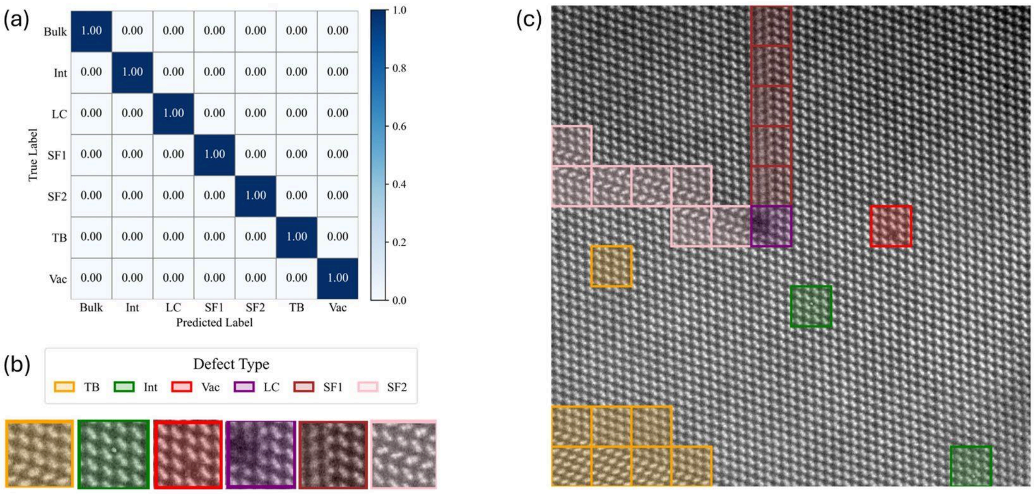


**Figure 1. Supervised defect classification in CdTe using the VGG-16 model. (a)** Normalized confusion matrix showing per-class prediction accuracy for the VGG-16-based defect classifier evaluated on the validation set. Perfect diagonal dominance and the absence of off-diagonal entries indicate that all seven image categories, Bulk, Interstitial (Int), Lomer Cottrell defect (LC), two different Stacking Faults (SF1, SF2), Twin Boundary (TB) and Vacancy (Vac) are classified without mislabeling on the validation data, demonstrating stable convergence and effective generalization of the model under the applied regularization and training strategy. **(b)** Classification of defect types and representative images of patch bearing corresponding defects. **(c)** Atomic images overlayed with the defect classification results.

To train the VGG16 model, the dataset consisted of 64×64 pixel grayscale STEM image patches representing seven different defect and bulk types. These patches were resized to 224×224 pixels and converted to three-channel inputs to meet the VGG16 input requirements, followed by standard ImageNet preprocessing. The trained model was then used to extract patch-level feature representations, enabling both supervised defect classification and unsupervised clustering of atomic-scale structures.

In the supervised setting, 144 patches of size 64×64 were extracted from a test image of size 768×768 pixel size image. These patches were classified into predefined region types, allowing identification of defect-bearing ROIs together with their spatial coordinates and confidence. Applied to CdTe (see Fig. 1), the framework identified 23 defect-related regions, substantially reducing the candidates for further multimodal analysis. Even with these cases, focusing on defect-bearing ROIs significantly improves analysis efficiency compared to exhaustive inspection.

Unsupervised analysis was performed on the VGG16-extracted patch embeddings using PCA for dimensionality reduction followed by K-means clustering, with the number of clusters determined by maximization of the silhouette score (see Fig. 2) [2]. Notably, the resulting seven-cluster solution showed strong correspondence with the supervised results, demonstrating that ROI identification is possible even in the absence of prior labels or explicit defect annotations.

Patch-level grouping further enables signal aggregation in spectroscopy or diffraction for dose-efficient defect analysis. In addition, confidence scores highlight patches that do not clearly belong to any category (confidence $< 0.4$), suggesting opportunities for the discovery of previously uncharacterized defect configurations. The approach was further validated on $SrTiO_3$ with artificial defects, indicating broad applicability and contributing to the accessibility of atomic-resolution STEM analysis by reducing reliance on expert-driven ROI selection.

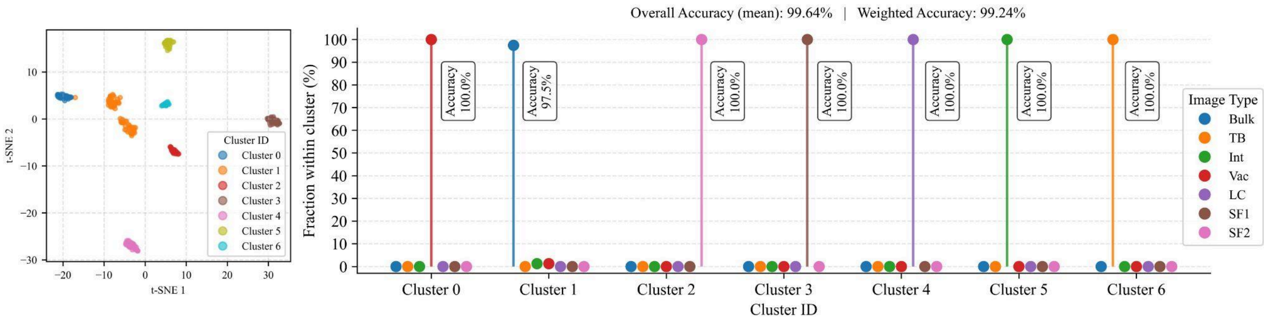


**Figure 2. Unsupervised clustering of CdTe defect embeddings using silhouette score maximization (k = 7).** Unsupervised clustering of CdTe STEM image embeddings using PCA followed by K-means clustering, where the number of clusters is selected based on the maximum silhouette-score (~0.72). The resulting k = 7 solution yields seven distinct clusters that show strong correspondence with the known CdTe defect classes, effectively capturing intra-class structural variations. Such selection of a silhouette optimum is a well-established and widely adopted practice in unsupervised machine learning. This higher-resolution clustering enables separation of defect types while preserving strong cluster compactness and inter-cluster separation, demonstrating that silhouette score can be used for physically meaningful clustering. The k = 7 solution is therefore selected for downstream analysis, achieving an overall clustering accuracy of 99.64%, with a sample-weighted accuracy of 99.24% as shown in the lollipop plot.

## Source of Data

https://github.com/RAW-Ayyubi/defect-classification-in-stem-images

The majority of the labeled image data were obtained from the above GitHub repository, where they were originally used by the authors for unsupervised clustering using convolutional variational autoencoders (CVAEs). In this work, we adopt the same dataset and repurpose it for VGG-16–based model construction, training, and evaluation.

## References:


[1] K. Simonyan, A. Zisserman, (214) arXiv:1409.1556.

[2] R. A. W. Ayyubi, S. Sultanov, J. P. Buban, R. F. Klie, [Undereview].