Atomic Lattice Disentanglement in STEM using rVAE

Jerry Zhang, Giovanni Romano, Hailey Reed
University of Connecticut

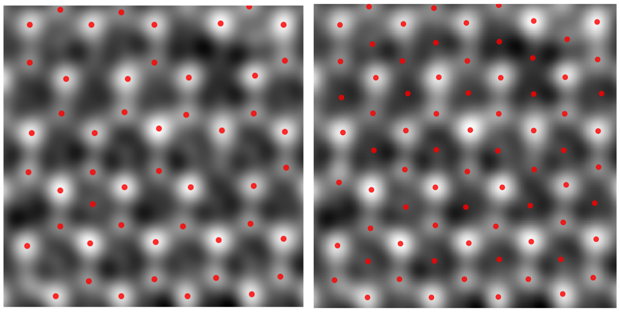
Abstract

Analyzing scanning transmission electron microscopy (STEM) images can be a time consuming task for experts. We would like to automate the task of identifying interesting features in these images but there do not exist large hand labeled dataset, due to the previous mentioned reason, to train conventional deep learning techniques on. In this project we seek to identify defects in MoS2 in an unsupervised fashion which would allow this method to be used circumnavigating the time consuming labeling process. To do this we propose using a rotationally invariant variational autoencoder to detect these anomalies and also provide a method to identify areas of interest to feed into the model.

Methods

To perform unsupervised defect detection in MoS2 we split the problem into two parts, determining where the lattice is and then extracting out patches where the atoms are or should be and then determining if there are any anomalies contained in the patch.

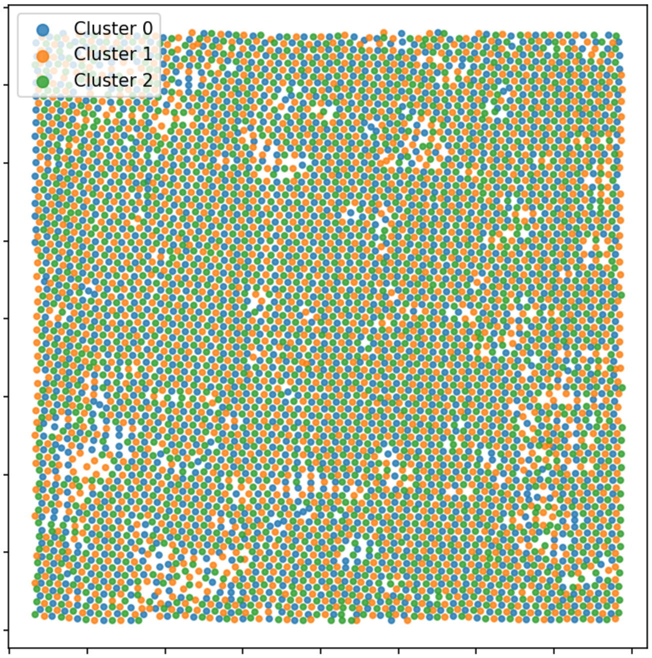
To determine where we should take our patches we take the bright areas of the image which generally correspond to atoms. First we preprocess the image using a bandpass filter to remove high and low frequency noise and then use peak detection to find local areas of brightness. While this gives us areas where atoms are located with high probability it does not capture areas where atoms should be but are not. To address this we leverage the fact that the atoms lie on a hexagonal lattice and attempt to fit a lattice to see where there should be points but we have not sampled from that region yet. Another concern is that image distortion will make it impossible to fit a single lattice over the whole image. This can be addressed by fitting a local lattice around each point and its neighbors as we can be relatively confident that this smaller region will lie on a lattice consistent with itself. We can see from the image that the method is able to pick points that may be too faint for just peak finding to locate.



After we obtain the patches we would like to be able to predict if a region is deviated from what is expected. To do this we employ a rotationally-invariant variational autoencoder (rVAE)[1] which is similar to a variational autoencoder (VAE) compressed data through a small latent space and reconstructs. The VAE also allows for a smooth latent space which allows for clusters to form in latent space which can be used to classify the data. The rVAE builds on top of the VAE by first predicting an angle to rotate the input image to obtain a canonical form, then passing it through the VAE to get the canonical reconstruction and finally rotates it back by its inverse to get the full reconstruction. This allows the model to explicitly remove rotation from being encoded into the latent space as defects are generally invariant to it. To predict the rotation we employ a spatial transformer network (STN) which is added to the VAE[2]. The STN is a small auxiliary network which predicts rotation using a differentiable matrix multiplication.

Results

To prepare our data we ran the peak finding + adaptive lattice method to create a dataset of patches then performed data augmentation by rotating, flipping and translation. To train the rVAE we first pretrained just the STN by feeding it a patch and a rotated image of that patch and forcing it to learn how to rotate the images. This was done in order to ensure the model learned how to rotate images constantly, then we initialized and froze the STN in the rVAE model and fully trained the rVAE model. For comparison we also trained a plain VAE on the same dataset.



To visualize results we clustered the latent representations using means and then plotted the cluster labels back onto the area where the image was taken. This clearly shows that the rVAE has learned on its own the difference between the molybdenum and sulfur atoms as well as the empty space. The next thing we looked at is determining if the model was able to find where there are defects. To do this we took each patch and took the patches with the worst reconstruction error as these represent rarer areas that the model was unable to generalize on. On the three patches at the bottom right of the page we clearly see that these patches that contain high reconstruction error contain defects either appearing to have molybdenum atoms in place of sulfur atoms or missing atoms all together

References

[1] Y. Liu, R. Proksch, C. Y. Wong, M. Ziatdinov, and S. V. Kalinin, “Disentangling Ferroelectric Wall Dynamics and Identification of Pinning Mechanisms via Deep Learning,” Advanced Materials, vol. 33, no. 43, p. 2103680, Oct. 2021, doi: 10.1002/adma.202103680.

[2] M. Jaderberg, K. Simonyan, A. Zisserman, and koray kavukcuoglu, “Spatial Transformer Networks,” in Advances in Neural Information Processing Systems, Curran Associates, Inc., 2015. Accessed: Dec. 17, 2025. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2015/hash/33ceb07bf4eeb3da587e268d663aba1a-Abstract.html