UNSUPERVISED CLASSIFICATION OF FERROELECTRIC DOMAINS

Grace Guinan, Arman Ter-Petrosyan,
Madeline Van Winkle, and Addie Salvador

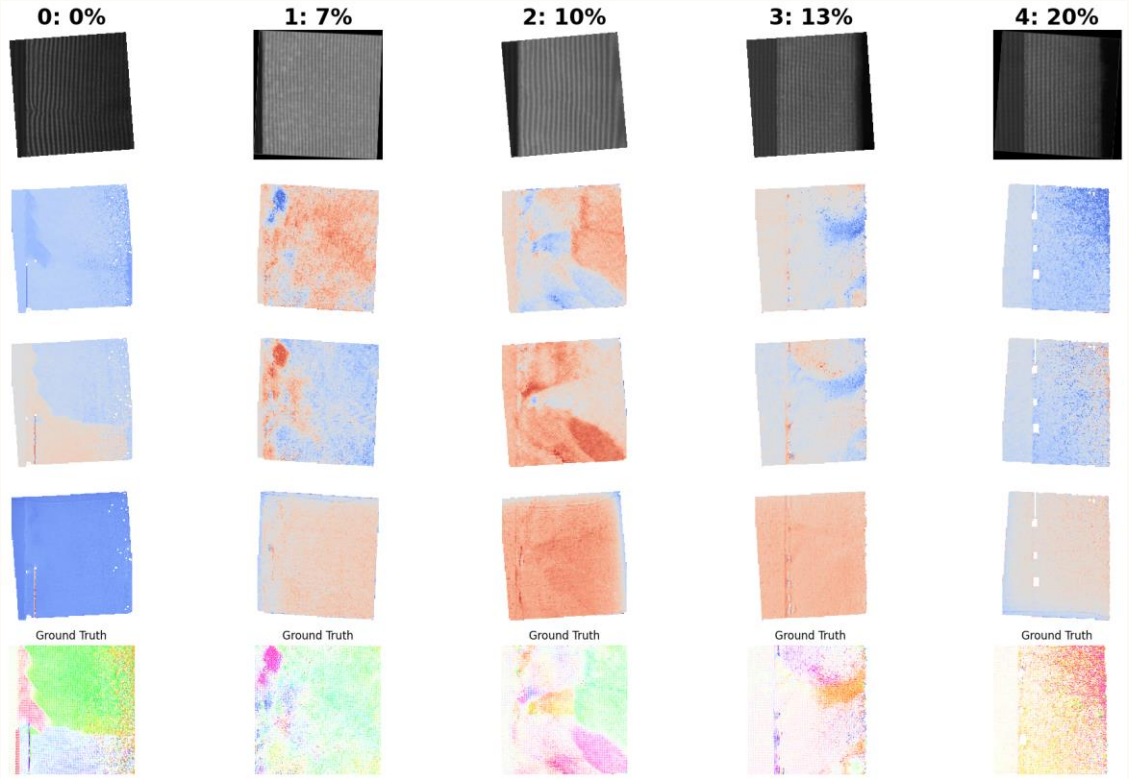
OBJECTIVE

Create a machine learning algorithm that can classify the different regions of a ferroelectric material autonomously

1) Optimize the patch window size
2) Test three different clustering methods
3) Align the labeling of the clusters across methods
4) Compare results

DATA SET

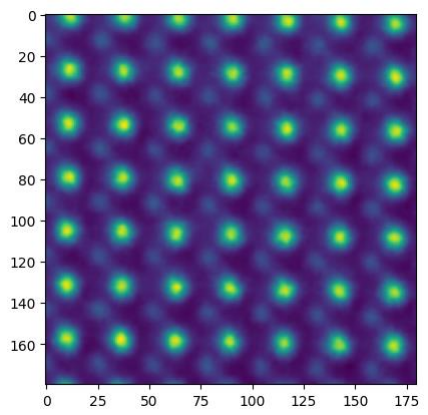
- A set of high-resolution STEM images of BiFeO3 doped with varying percentages of Sm were given
- For this algorithm, only the pure BFO was analyzed since it has the most defined regions



These images were obtained by Dr. Chris Nelson of Oak Ridge from samples prepared by Professor Ichiro Takeuchi at the University of Maryland.

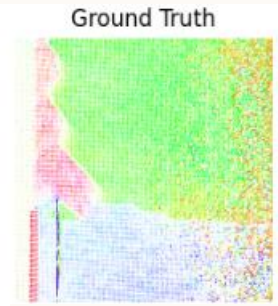
OPTIMIZING WINDOW SIZE

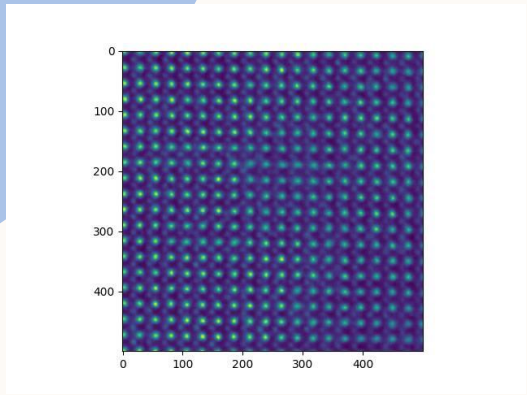
- Three parameters to adjust:
  - Length in x
  - Length in y
  - Step size
- Started with a non-overlapping square chip (180,180) to get cluster methods running and then used an iterative approach to optimize the three parameters

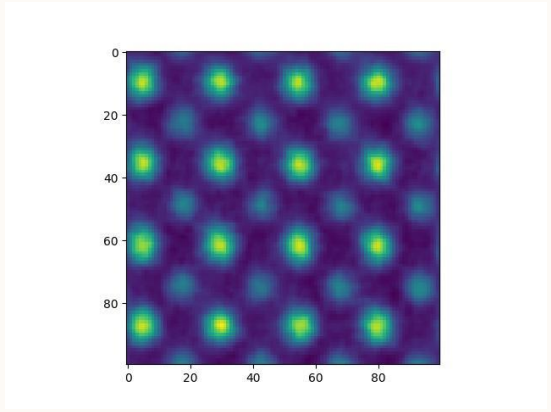


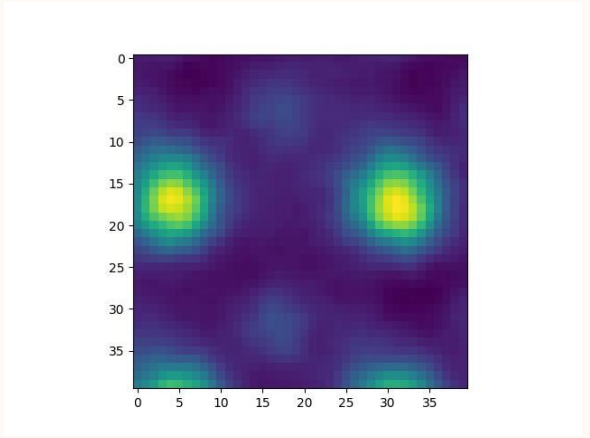

```
window_size = (180,180)

step_size = 180

imstack_grid, com_grid = custom_subimages(imgdata[0], step_size, window_size)
```

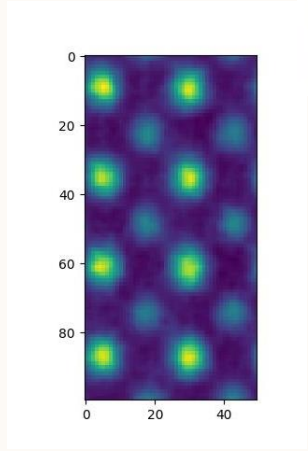
K-MEANS: WINDOW SIZE

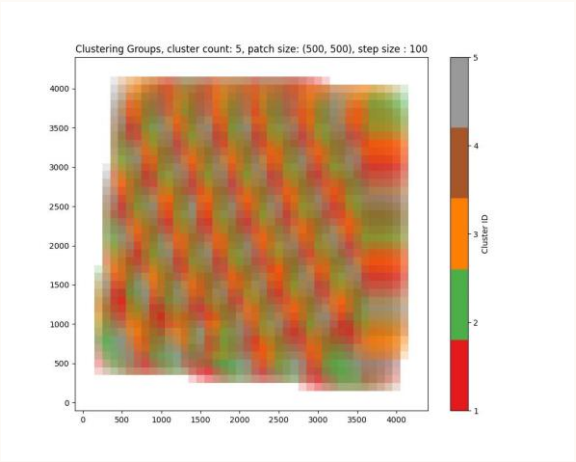


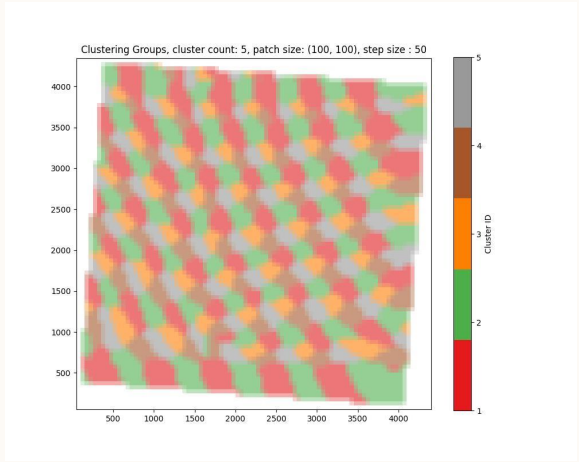


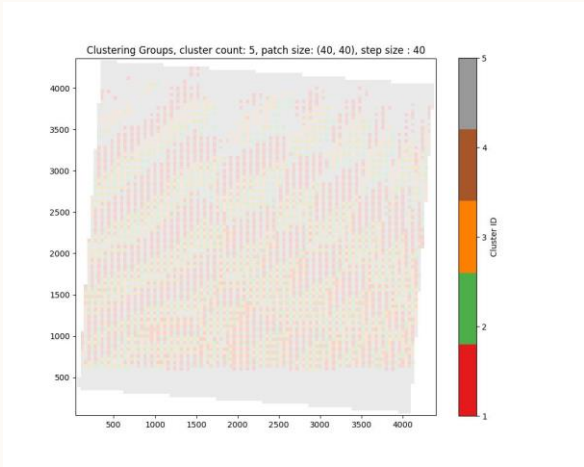


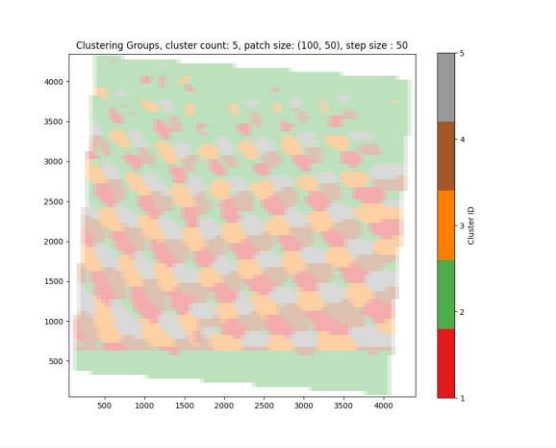










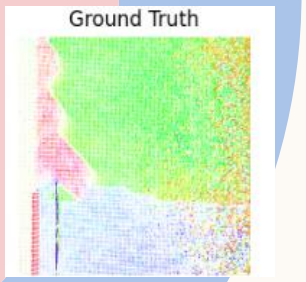


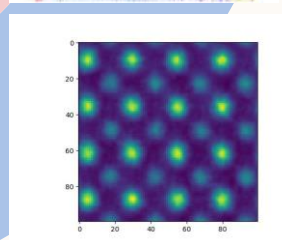
Clusters: 5
Window size: 500, 500
Step size: 100

Clusters: 5
Window size: 100, 100
Step size: 50

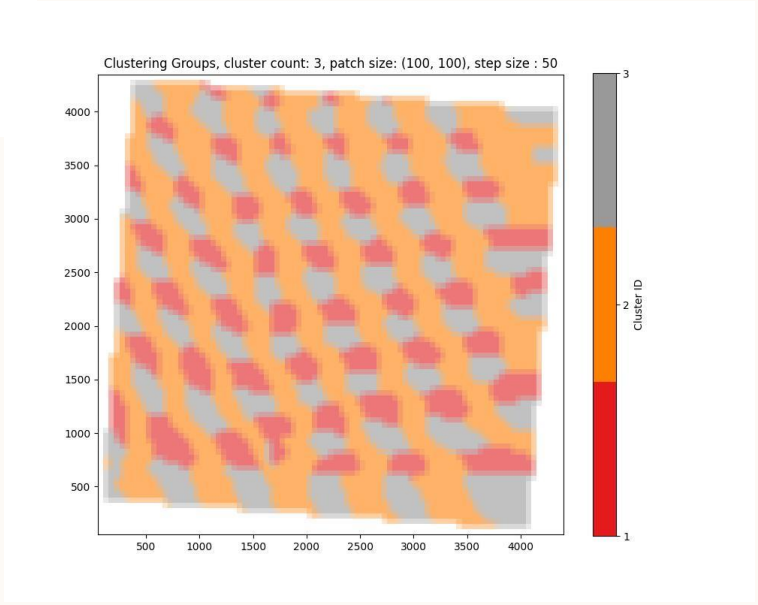
Clusters: 5
Window size: 40, 40
Step size: 40

Clusters: 5
Window size: 100, 50
Step size: 50

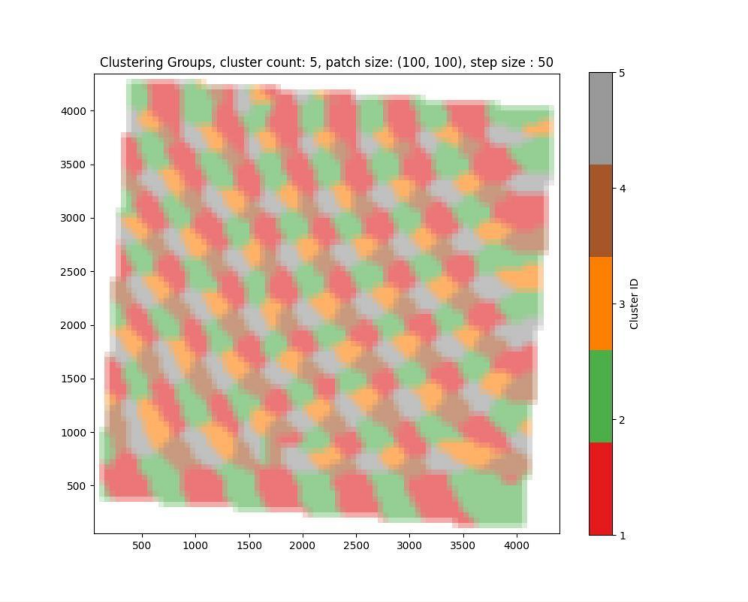




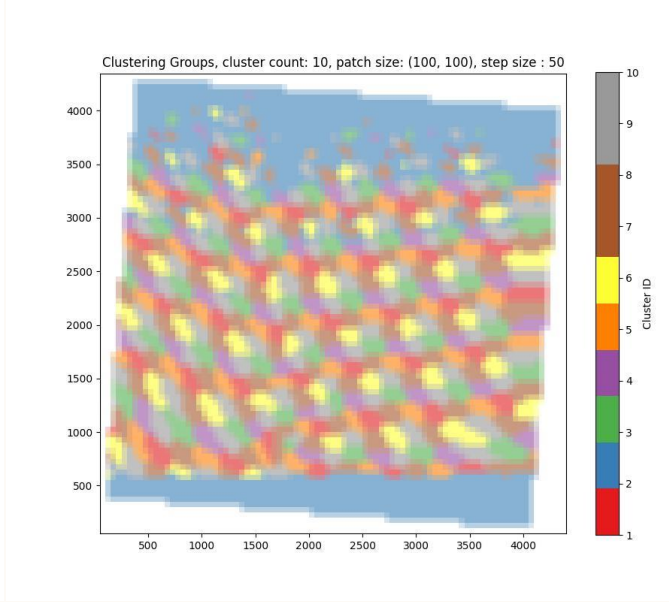
K-MEANS: NUMBER OF CLUSTERS



Clusters: 3
Window size: 100, 100
Step size: 50

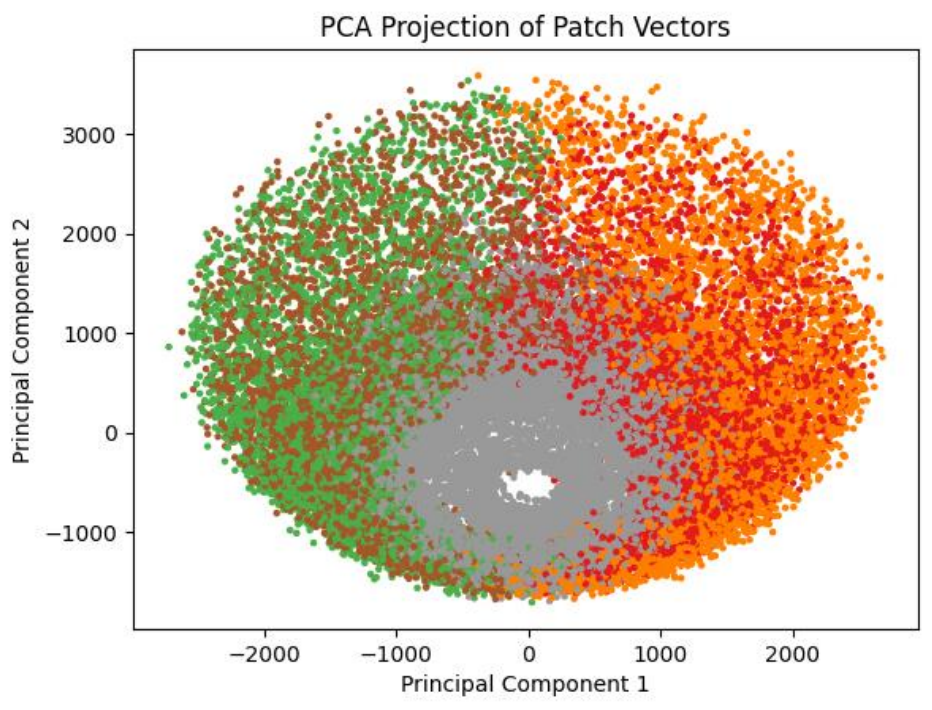


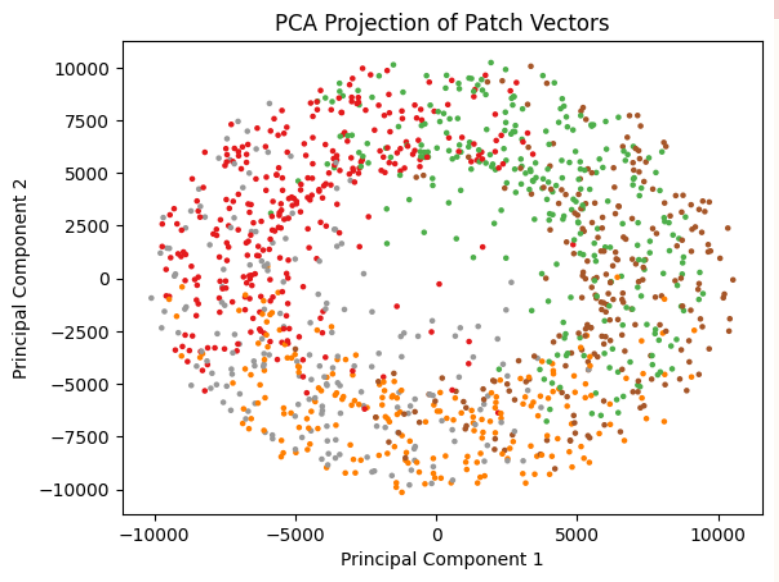
Clusters: 5
Window size: 100, 100
Step size: 50

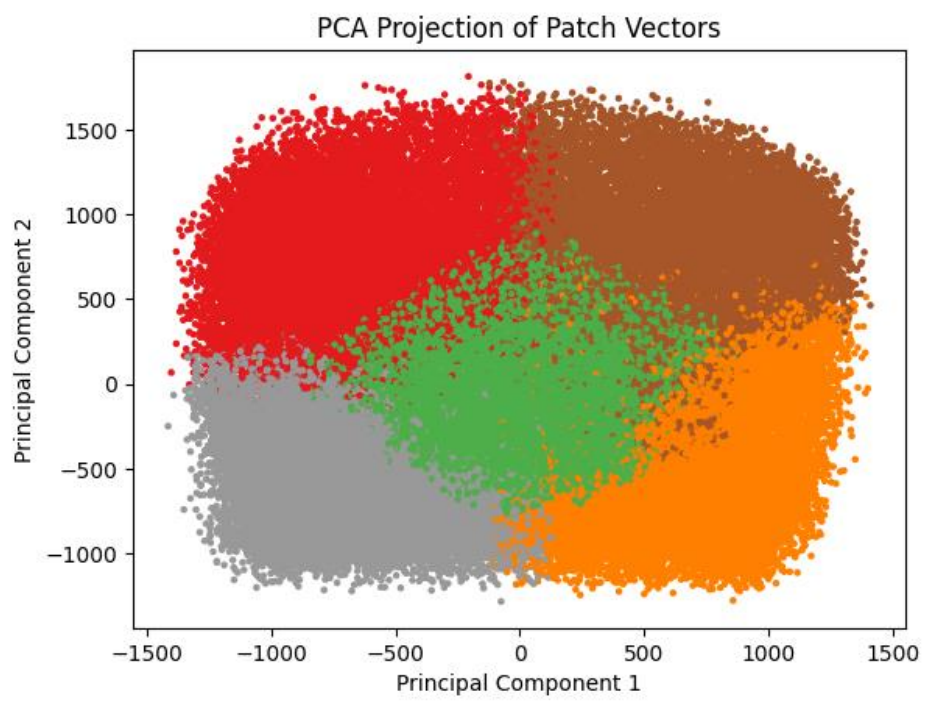


Clusters: 10
Window size: 100, 100
Step size: 50

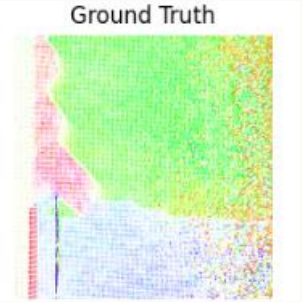
PCA ON DATA


PCA Projection of Patch Vectors
Principal Component 2
Principal Component 1
3000
2000
1000
0
−1000
−2000
−1000
0
1000
2000

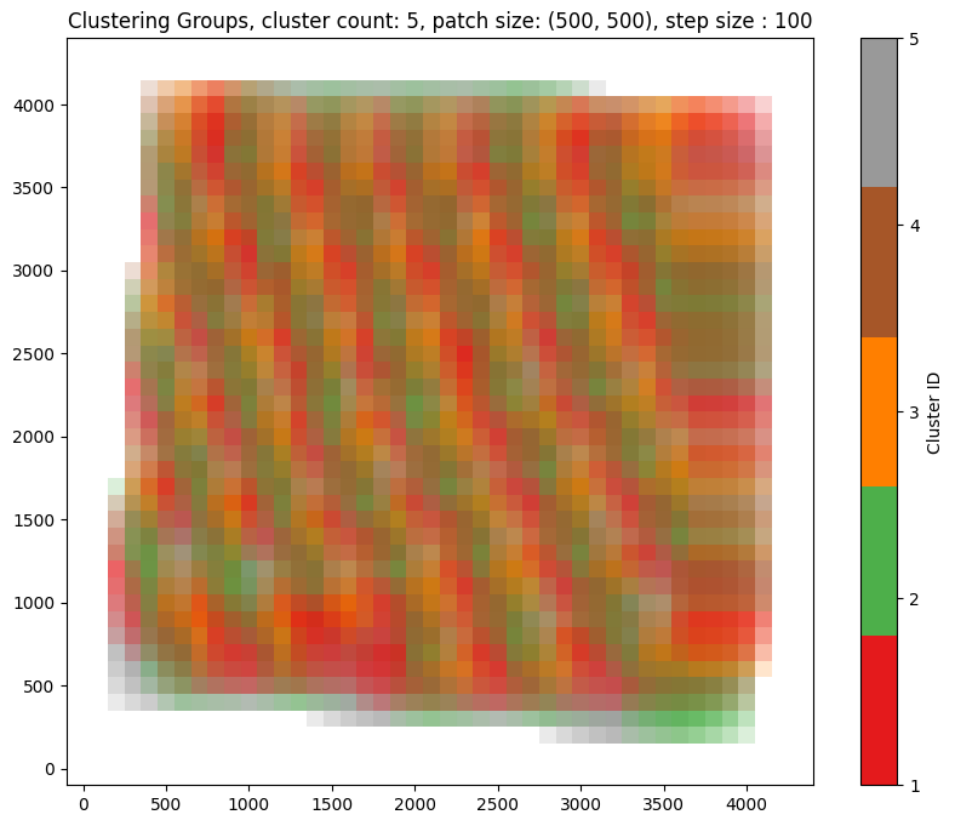

PCA Projection of Patch Vectors
Principal Component 2
Principal Component 1
10000
7500
5000
2500
0
−2500
−5000
−7500
−10000
−10000
−5000
0
5000
10000


PCA Projection of Patch Vectors
Principal Component 2
Principal Component 1
1500
1000
500
0
−500
−1000
−1500
−1000
−500
0
500
1000
1500

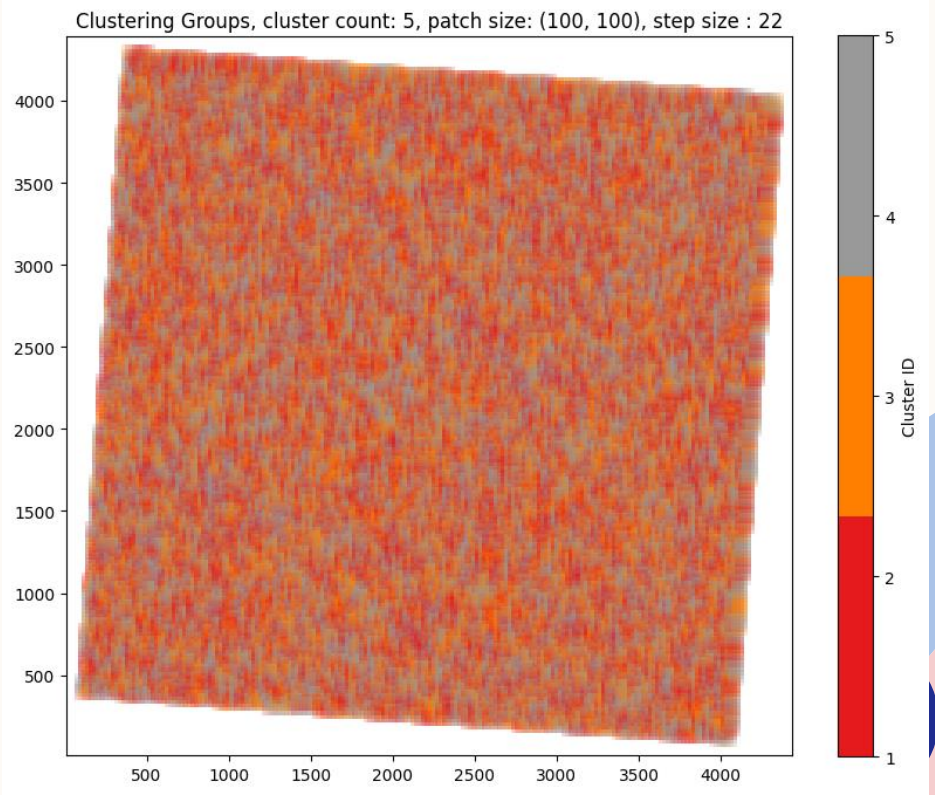
RESULTS FROM AGGLOMERATIVE AND SPECTRAL



Agglomerative



Spectral



CONCLUSIONS

- Learned how a variety of clustering methods work
- Learned how to collaborate on a coding project

Next steps:

- From PCA visualization, try a circular clustering method
- Embed data before clustering (using imagenet or micronet)
- To compare across clustering methods – align data by which clusters had the most overlapping tiles

- Maybe try a non-clustering method (locate atomic coordinates? Or supervised learning (few-shot)?)

THANK YOU