

PNNL-34214

Use of Graph Theory and Neural Networks for Microstructural Classification

April 2023

- 1 Arman H. Ter-Petrosyan
- 2 Jenna Pope
- 3 Christina M. Doty
- 4 Bethany E. Matthews
- 5 Sarah M. Akers
- 6 Steven R. Spurgeon

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor Battelle Memorial Institute, nor any of their employees, makes **any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights.** Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or Battelle Memorial Institute. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from
the Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062

www.osti.gov
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
or (703) 605-6000
email: info@ntis.gov
Online ordering: <http://www.ntis.gov>

Use of Graph Theory and Neural Networks for Microstructural Classification

April 2023

1 Arman H. Ter-Petrosyan
2 Jenna Pope
3 Christina M. Doty
4 Bethany E. Matthews
5 Sarah M. Akers
6 Steven R. Spurgeon

Prepared for
the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99354

Abstract

Recent advances in materials data analytics have provided new avenues for determining process-structure-property (PSP) linkages in a variety of materials. Machine learning techniques including few-shot learning have increased the efficiency of classifying microscopy images for the purposes of material characterization. Modifications in segmentation also show potential in improving the accuracy of our current pyCHIP classifier. Replacing previous encoders trained on ImageNet with those trained on microscopy images like MicroNet has initially shown better performance at classifying images of irradiated samples. Additionally, different normalization approaches were tested to show no discernable effect on classification. The Louvain method for community detection is analyzed on a set of irradiated samples with different parameters to determine which proved beneficial under what circumstances. We suggest that microscopy experiments be automated in the future using a combination of these techniques to enable high-throughput analyses.

Acknowledgments

A.H.T. was supported in part by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internships Program (SULI). This research was supported by a Chemical Dynamics Initiative (CDi) Laboratory Directed Research and Development (LDRD) project at Pacific Northwest National Laboratory (PNNL). PNNL is a multiprogram national laboratory operated for the U.S. Department of Energy (DOE) by Battelle Memorial Institute under Contract No. DE-AC05-76RL0-1830.

A big thank you to Steven Spurgeon, Jenna Pope, Christina Doty, Bethany Matthews, and Sarah Akers at PNNL. Steven was gracious enough to make time in his busy schedule to work through any problems I came across throughout my internship, ensuring that my research experience was both productive and informative. Jenna introduced me to multiple software packages and regularly helped me write code for my project. Christina introduced me to the research team's codebase, helped me setup my computing environment for the project, and addressed any issues regarding version control. Bethany provided image data and introduced me to Pathfinder. Sarah addressed questions I had regarding the multimodal model discussed in the paper.

Contents

Abstract	ii
Acknowledgments	iii
1.0 Introduction	1
2.0 Results and Discussion	2
2.1 Supervised Classification.....	2
2.2 Unsupervised classification	2
2.3 Network Graph Analysis	3
2.4 Reproducibility and Community Analysis	4
3.0 Conclusions.....	7

Figures

Figure 1. The image on the left (a) is an image of a SrTiO₃/Ge alloy taken by a STEM microscope. Note the different regions of the materials microstructure as well as the interfaces between them. On the right (b) is an overlay image showing the

different regions in the image as picked up by the classifier. While it does not do a perfect job at classifying the different regions present in the image, it gets most. (2)

Figure 2. (From left to right) (a) An original image of a pristine metal oxide sample. (b) The corresponding graph with Louvain communities colored. Each node represents a chip from the original image and chips which fall under the same community are assigned the same color in the graph. The communities result from a cutoff value of 0.9 for the Louvain method (c) An overlay image which visually maps the chips and their assigned community to their location in the original image. Note how the unsupervised classifier does a good job at classifying the different regions seen in the original image. (4)

Figure 3. (From left to right, up and down) (a) (c) An original repeat image of the same pristine metal oxide sample shown in Figure 2. (b) (d) An overlay image which visually maps the chips and their assigned community to their location in the original image. From looking at the overlay images in comparison to the originals, we conclude that the methods work as expected since they appear to pick up on several of the characteristics a domain expert would see when analyzing these images. (5)

Figure 4. Corresponding histograms that show the distributions of pixel intensities for all chips grouped by community. Note that communities with only one member are not shown. (From right to left) (a) The first repeat image (Figure 3 (a) (b)). (b) The second repeat image (Figure 3 (c) (d)). These histograms (along with the overlay images in Figure 3) illuminate some interesting features of the current methods used for classification in our research. Notice that although the distributions of communities 1 and 2 as well as 3 and 4 seem to overlap in the bottom histogram, the communities are still well separated in the overlay image. This is not the case in the top histogram where communities 1 and 4 seem to overlap but are not well separated in the overlay image. It appears that there are some features that the Louvain method is picking up on in one instance but not the other. (6)

1.0 Introduction

A longstanding objective of materials science research is the discovery of process-structure-property (PSP) linkages. The application of data science techniques to materials data allows for increased efficiency in mining and extracting valuable materials knowledge. (Kalidindi & De Graef, 2015) This materials knowledge can accelerate the development of novel materials for a variety of modern applications including energy storage, quantum computing, and transformative manufacturing. (Olszta, et al., 2022)

One possible avenue is autonomous experimentation. Integrating artificial intelligence (AI) and machine learning (ML) into current microscopy setups allows researchers to conduct high-throughput analyses. Olszta et al. demonstrated the use of few-shot ML feature classification in a scanning transmission electron microscope (STEM) to automatically identify desired features in microscopy image data depending on the task at hand. (Olszta, et al., 2022) Classification algorithms can be used to analyze the microstructure of materials. Akers et al. have developed a few-shot learning implementation that offers experimenters the ability to both rapidly and accurately classify images taken by microscopes using a small training set. (Akers, et al., 2021) The pyCHIP classifier is one such model that implements this functionality and has proven to be useful in classifying images of several different kinds of materials. (Doty, et al., 2022) The development of unsupervised ML methods for classifying image data of metal oxides will be discussed at length in this paper as an additional approach to the current pyCHIP classifier. Incorporating the results of both the supervised and unsupervised ML methods will help establish trends in analyzing microscopy data which can be exploited for further development of classification models.

2.0 Results and Discussion

The current iteration of the pyCHIP classifier segments images into square chips of equal dimension, allowing researchers to easily define support sets. Though straightforward, this approach has proven to be useful at accurately classifying the composition of a material's microstructure. Other segmentation techniques can provide additional benefits besides improved accuracy such as feature detection. Specifically, unsupervised ML models have been used for the segmentation of static and dynamic atomic-resolution microscopy data sets in the form of images and video sequences. (Wang, Freysoldt, Zhang, Liebscher, & Neugebauer, 2021) Models like the Louvain method for community detection help detect similarities in data and are simple to implement. Many network graphing packages include clustering or community detection capabilities and are great for visualization purposes.

2.1 Supervised Classification

The pyCHIP classifier is a few-shot learning model and thus only requires a small training set to accurately classify an image. To define a training set, the image is first segmented into a grid of smaller square images or chips of equal dimension. The researcher can then pick individual chips as representatives for each class of particle present in the image. These representative chips define the support set for the class. Once support sets for each class are defined, the model classifies the remaining chips in the image by comparing them to the support sets. The classifier output is a data frame containing the class probabilities for each chip as well as an overlay image like that in Figure 1.

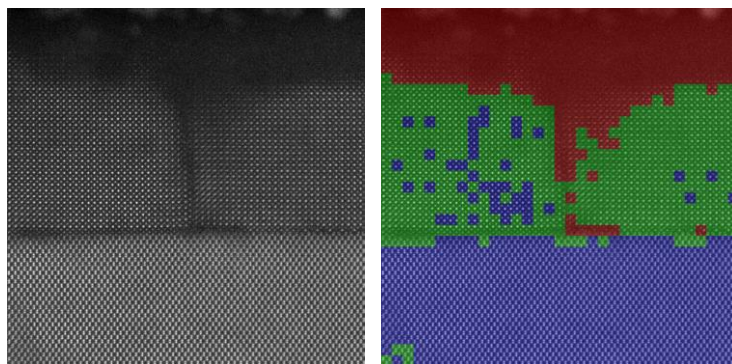


Figure 1. The image on the left (a) is an image of a SrTiO₃/Ge alloy taken by a STEM microscope. Note the different regions of the materials microstructure as well as the interfaces between them. On the right (b) is an overlay image showing the different regions in the image as picked up by the classifier. While it does not do a perfect job at classifying the different regions present in the image, it gets most.

2.2 Unsupervised classification

While the supervised approach does a decent job, it can still be improved upon. One such improvement is to remove the researcher entirely from the decision-making process. This leads us to the development of an unsupervised approach to the problem of classifying images.

In the unsupervised classifier, the model has no access support sets and will have to discriminate the chips to form groups or clusters of chips with similar features. The model's performance can then be compared to the results of the supervised approach. To further increase the model's discriminating power, a method was defined to quantify the similarity amongst chips by calculating their distances to other chips and return the results in data frame. A PychipClassifier method, chip2chip accepts several parameters: dataset, encoder, maximum query size, seed, save path, filename, and cutoff. The dataset parameter can be set to either "ImageNet" (default) or "MicroNet". (MicroNet is a NASA database of microscopy images. As one might suspect, the use of an encoder trained on microscopy images is preferable in this context. The qualitative results of the latest analysis presented in this report prove this to be the case.) The encoder parameter can be set to one of seven PyTorch encoders used by the pyCHIP classifier: resnet18, resnet34, resnet50, resnet101, resnet152, torch101, and shufflenet. Note that the MicroNet dataset is currently available with only a resnet50 encoder.

The cutoff parameter can be set to any positive rational value and represents the acceptable upper limit for... Thus, chips with classifier output less than the cutoff are used when computing the distances.

Once the similarity is computed, the data is organized in a data frame with columns representing the chips, distances, and normalized distances. A summary of the steps implemented by the chip2chip method is given below.

1. Load the pretrained encoder
2. Chip the image
3. Obtain encodings of each chip
4. Compute the distance between each pair of chips
5. Return the data in a dataframe

2.3 Network Graph Analysis

Using the output of chip2chip, a network graph is constructed using Python's NetworkX package. The nodes of the network are the chips and edges are drawn connecting all nodes. Since shorter distances correspond to greater similarity and farther distances correspond to less similarity, the normalized distances are subtracted from one to provide the appropriate weights used for building the network graph. An example of a network graph with the Louvain method used to group chips into communities of similar chips is shown in the figure below.

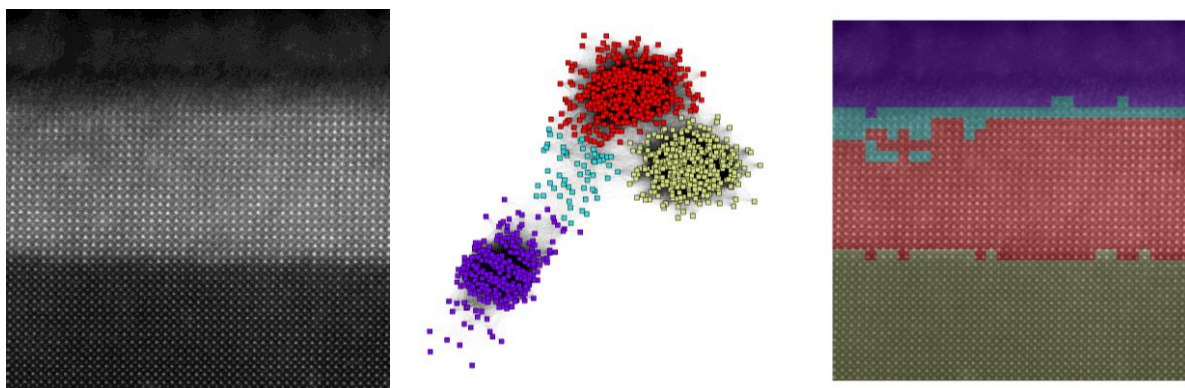


Figure 2. (From left to right) (a) An original image of a pristine metal oxide sample. (b) The corresponding graph with Louvain communities colored. Each node represents a chip from the original image and chips which fall under the same community are assigned the same color in the graph. The communities result from a cutoff value of 0.9 for the Louvain method (c) An overlay image which visually maps the chips and their assigned community to their location in the original image. Note how the unsupervised classifier does a good job at classifying the different regions seen in the original image.

When using the Louvain method, it is necessary to specify a cutoff value for the graph's weights. Multiple values for the cutoff were tested to determine which yielded the best results for a series of images. This allows us to determine when a particular cutoff picks up on any discernable features in an image.

2.4 Reproducibility and Community Analysis

It is always good idea when developing a new series of ML tools to use in analyzing a broad array of data to ensure that the methods used are in fact reproducible on similar data. In this vein, we ran the same analysis as before on a series of repeat images for the metal oxide sample in Figure 2.

Minor changes were made during this step including normalizing the distances across all images in the series as opposed to a per image basis as well as removing the final layer in the encoders used. In addition to qualitatively analyzing the performance of our unsupervised classifier, we decided to look at the distribution of pixel intensities across chips in an image grouped based on their community membership. The results of two such repeat images are shown in the figures below.

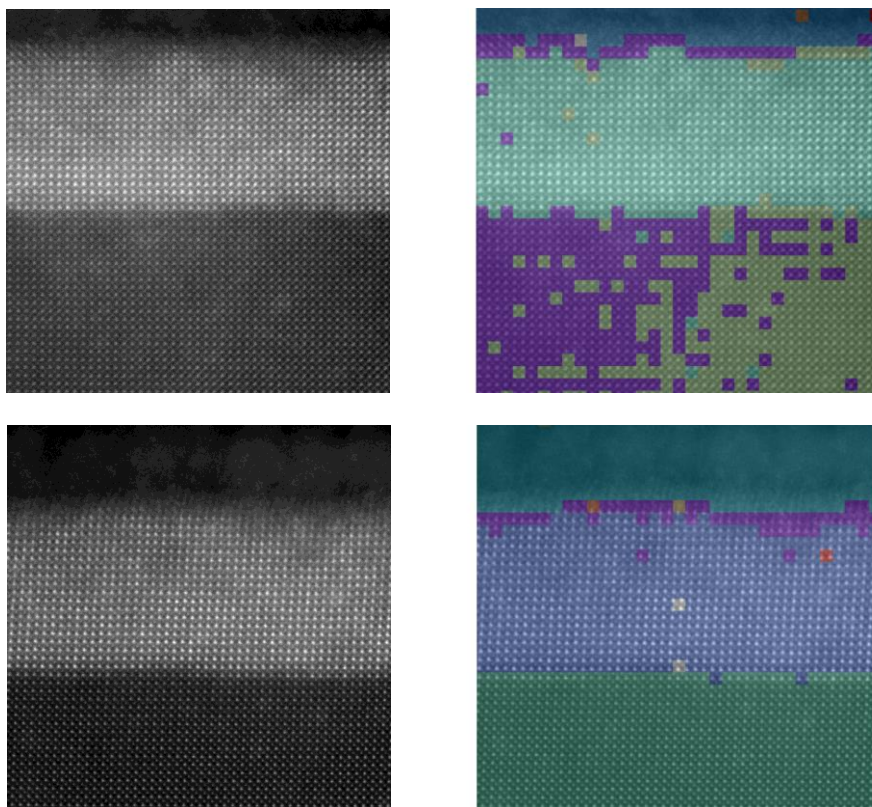


Figure 3. (From left to right, up and down) (a) (c) An original repeat image of the same pristine metal oxide sample shown in Figure 2. (b) (d) An overlay image which visually maps the chips and their assigned community to their location in the original image. From looking at the overlay images in comparison to the originals, we conclude that the methods work as expected since they appear to pick up on several of the characteristics a domain expert would see when analyzing these images.

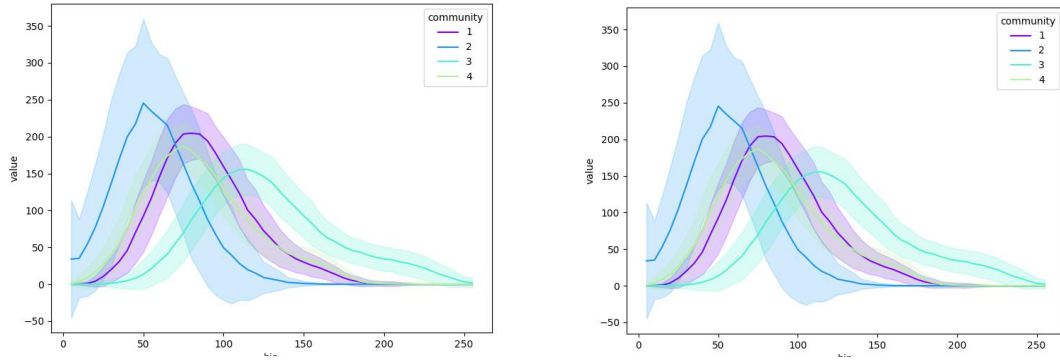


Figure 4. Corresponding histograms that show the distributions of pixel intensities for all chips grouped by community. Note that communities with only one member are not shown. (From right to left) (a) The first repeat image (Figure 3 (a) (b)). (b) The second repeat image (Figure 3 (c) (d)). These histograms (along with the overlay images in Figure 3) illuminate some interesting features of the current methods used for classification in our research. Notice that although the distributions of communities 1 and 2 as well as 3 and 4 seem to overlap in the bottom histogram, the communities are still well separated in the overlay image. This is not the case in the top histogram where communities 1 and 4 seem to overlap but are not well separated in the overlay image. It appears that there are some features that the Louvain method is picking up on in one instance but not the other.

3.0 Conclusions

While the current iteration of the pyCHIP classifier allows researchers to easily classify microscopy images, additional improvements can be made to fully automate the experimentation process. Further research on unsupervised approaches can eventually lead to the removal of the researcher from the decision-making process and thus enable high-throughput analyses. Expanding upon this research and studying the use of megagraphs (graphs constructed with chips from all images in a dataset, not just a singular image) is a logical next step. This coupled with the research being done by other members of our group will help reveal new insights to develop tools for the material science community.

References

- Akers, S., Kautz, E., Trevino-Gavito, A., Olszta, M., Matthews, B. E., Wang, L., . . . Spurgeon, S. R. (2021, November 17). Rapid and flexible segmentation of electron microscopy data using few-shot machine learning. *npj Computational Materials*. Retrieved from <https://www.nature.com/articles/s41524-021-00652-z>
- Doty, C., Gallagher, S., Cui, W., Chen, W., Bhushan, S., Oostrom, M., . . . Spurgeon, S. R. (2022, February 15). Design of a graphical user interface for few-shot machine learning classification of electron microscopy data. *Computational Materials Science*, 203. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0927025621007874>
- Kalidindi, S. R., & De Graef, M. (2015). Materials Data Science: Current Status and Fututre Outlook. *Annual Review of Materials Research*, 45, 171-193.
- Olszta, M., Hopkins, D., Fiedler, K. R., Oostrom, M., Akers, S., & Spurgeon, S. R. (2022, June 10). An Automated Scanning Transmission Electron Microscope Guided by Sparse Data Analytics. *Microscopy and Microanalysis*, 28(5). Retrieved from <https://www.cambridge.org/core/journals/microscopy-and-microanalysis/article/an-automated-scanning-transmission-electron-microscope-guided-by-sparse-data-analytics/2925BA91037B33A7523ABFEEEB39F6DA>
- Wang, N., Freysoldt, C., Zhang, S., Liebscher, C. H., & Neugebauer, J. (2021, September 1). Segmentation of Static and Dynamic Atomic-Resolution Microscopy Data Sets with Unsupervised Machine Learning Using Local Symmetry Descriptors. *Microscopy and Microanalysis*, 1-11.

Pacific Northwest National Laboratory

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99354

1-888-375-PNNL (7665)

www.pnnl.gov