

High-Throughput Identification of Crystal Structures Via Machine Learning

Kevin Kaufmann¹, Chaoyi Zhu², Alexander S. Rosengarten¹, Daniel Maryanovsky³, Tyler Harrington², Eduardo Marin¹ and Kenneth Vecchio^{1,2*}

¹. Department of NanoEngineering, UC San Diego, La Jolla, CA, USA.

². Materials Science and Engineering Program, UC San Diego, La Jolla, CA, USA.

³. Department of Cognitive Science, UC San Diego, La Jolla, CA, USA.

* Corresponding author: kvecchio@eng.ucsd.edu

Identifying crystal structure is a crucial step in the analysis of proteins, micro- and macro-molecules, pharmaceuticals, geological specimens, and synthetic materials [1-3]. The most common practices involve analysis of diffraction patterns produced in laboratory X-ray diffractometers, transmission electron microscopes, and synchrotron X-ray sources. However, these techniques are slow, require careful sample preparation, can be difficult to access, and are prone to human error during analysis. Traditional electron backscatter diffraction (EBSD) is comparatively faster and easier to perform; however, contains the caveat that you have already determined the phases in your sample.

This work presents a newly developed methodology that represents a paradigm change in electron diffraction-based structure analysis techniques, with the potential to revolutionize multiple crystallography-related fields. A machine learning-based approach for rapid and autonomous identification of the crystal structure of metals and alloys, ceramics, and geological specimens, without any prior knowledge of the sample, is demonstrated utilizing the electron backscatter diffraction technique. This methodology can be utilized to assist in identifying an unknown phase or produce phase maps of the crystal structure at each location (see Fig. 1 for the methodology of this machine learning approach).

The applicability of this approach is evaluated on diffraction patterns from samples unknown to the computer without any human input or data filtering. Furthermore, the internal operations of the neural network are elucidated through visualizing the symmetry features learned by the convolutional neural network and it is determined that the model looks for the same features a crystallographer would use, even though it is not explicitly programmed to do so. Instead, the deep learning system takes in raw data and automatically discovers the representations that maximize classification performance [4]. Fig. 2 shows examples of the regions in the EBSD patterns where deep learning was most strongly activated: zone axes. This study opens the door to fully automated, high-throughput determination of crystal structures via several electron-based diffraction techniques. Convergent beam electron diffraction (CBED) is an excellent candidate for transfer learning of this model since it also utilizes Kikuchi diffraction lines. Moreover, the deep learning methodology employed herein could be applied to other disciplines outside of materials science including geology, pharmacology, and even structural biology [5].

References:

[1] T Walz et al., *Nature* **387** (1997), p. 624.

[2] S Curtarolo et al., *Phys. Rev. Lett.* **91** (2003), p. 135503.

[3] JM Cowley et al., International Tables for Crystallography (2006), p. 276.

doi:10.1107/97809553602060000558

[4] Y LeCun, Y Bengio, and G Hinton, Nature **521** (2015), p. 436.

[5] K. Kaufmann was supported by the Department of Defense (DoD) through the National Defense Science and Engineering Graduate Fellowship (NDSEG) Program. K. Kaufmann would also like to acknowledge the support of the ARCS Foundation.

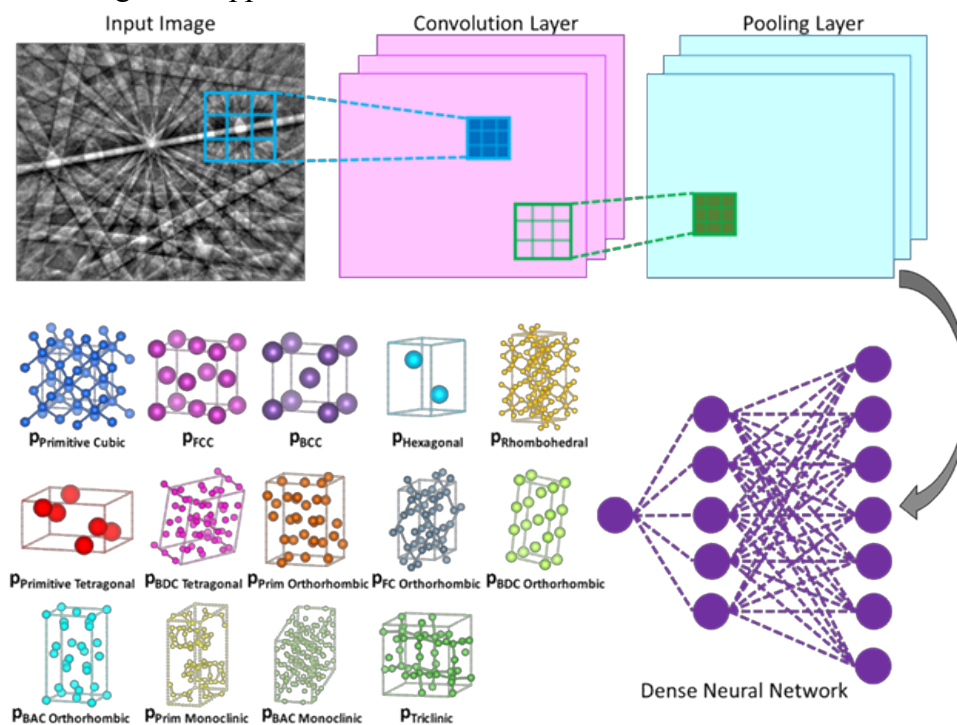


Figure 1. Illustration of the inner workings of a convolutional neural network. Each convolutional layer extracts features from its preceding layer, using filters (or kernels) learned from training the model, to form feature maps. These feature maps are then down-sampled by a pooling layer to exploit data locality. A traditional dense neural network, a simple type of classification network, is placed as the last layer of the CNN, where the probability that the input diffraction pattern belongs to a given class (e.g. Bravais lattice or point group) is computed.

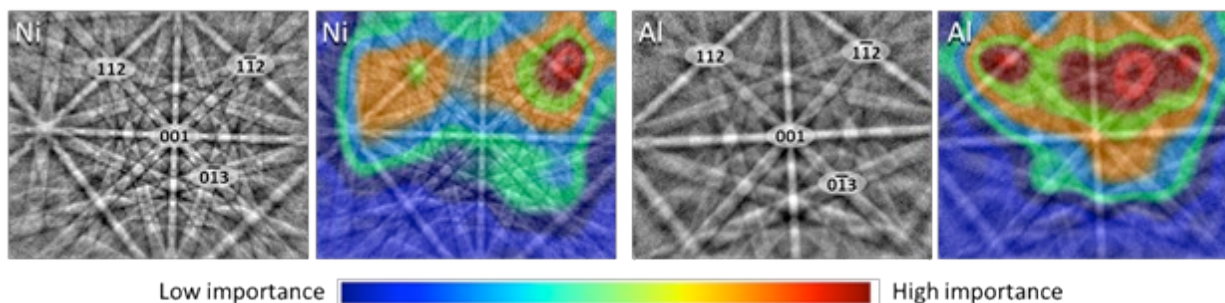


Figure 2. Visualizing the features utilized for classification of a diffraction pattern to face-centered cubic (FCC). An electron backscatter diffraction pattern from nickel (Ni) and aluminum (Al) were selected from nearly identical orientations. The corresponding heatmap displays the normalized importance of information in the image for correctly classifying it as FCC. It is observed that for each of these two images, the symmetry information near the $[1-12]$ zone axis produces the highest activation, followed by the $[112]$ zone axis, and symmetry shared by the $[001]$ and $[0-13]$ zone axes.