Materials Discovery: Understanding Polycrystals from Large-Scale Electron Patterns

3rd Workshop on Advances in Software and Hardware (ASH) at BigData
December 5, 2016

Ruoqian Liu, Dipendra Jha, Ankit Agrawal, Wei-keng Liao, Marc De Graef and Alok Choudhary
Hypothesis:

We can make use of **big data + deep learning + advanced hardwares and softwares** to make interesting discoveries - in materials!
Outline

• Deep Learning: the Basics
• Materials Discovery: the Basics
• Deep CNN for EBSD Indexing
• Advanced Software and Hardware in Use
Outline

• **Deep Learning: the Basics**
  • Materials Discovery: the Basics
  • Deep CNN for EBSD Indexing
  • Advanced Software and Hardware in Use
What is Deep Learning

• A revolutionary breakthrough in machine learning

• Caused big companies, such as Microsoft, Facebook, Google, Apple, Baidu Yahoo! and IBM to heavily invest in this technology

• The perfect method to build large-scale recognition systems to exploit the information locked away in Big Data

• Boosted new surge of Artificial Intelligence
What Does Deep Learning Learn

Learning the representation

• The way in which data are represented can make a huge difference in the success of a learning algorithm.

• Deep learning enables the learning of multiple levels of representation, discovering more abstract features in the higher levels.

Learning as human does

• Because human brains appear deep, AI-tasks require deep circuits

• Because it is natural for humans to represent concepts at multiple levels of abstractions, deep architecture makes sense.

• Because human learn mostly unsupervised, only partially supervised.
Three Keys in Deep Learning

- Deep neural networks
- Big data
- Fast computing
DL Achievements in Various Fields

- **Computer vision**: where DL first showed its power.
- **Speech recognition**: improved by 30% — an earthquake in this field.
- Genetics, drug discovery, health, ...
- Materials science
Convolutional Neural Networks

- Convolutional Neural Networks (ConvNets or CNNs) are a category of Neural Networks that have proven very effective in learning image tasks.

Given images, produce relevant tags

Given images, recognize/locate objects
CNN for Image Classification

- CNNs can come with different architecture. Below is similar to what’s known as the LeNet. It classifies an input image into four categories: dog, cat, boat or bird.

On receiving a boat image as input, the network correctly assigns the highest probability for boat (0.94) among all four categories. The sum of all probabilities in the output layer should be one.

Key Operations in a CNN

There are 4 main operations in the CNN LeNet.

1. Convolution
2. Non Linearity (ReLU)
3. Pooling or Sub Sampling
4. Classification (Fully Connected Layer)

First of all, images are a matrix of pixel values
- A colored image will have three channels – red, green and blue
- A grayscale image has just one channel.
Outline

- Deep Learning: the Basics
- **Materials Discovery: the Basics**
- Deep CNN for EBSD Indexing
- Advanced Software and Hardware in Use
Materials Discovery: A Fourth Paradigm

- Empirical Science: Laboratory Experiments
  - Laws of Thermodynamics

- Theoretical Science
  - Simulations: Density Functional Theory, Molecular Dynamics

- Computational Science
  - Predictive analysis, Relationship mining, Anomaly detection

- Data Science

1600 1950 2000


Deep Learning for Materials Science Applications

Fast computing?  Yes! (good enough)
Big data?  Materials Project:

https://www.materialsproject.org
Open Quantum Materials Database:

http://www.oqmd.org
Deep neural networks?  We are the first!
Our specific problem in materials discovery:
Electron imaging (observation) → polycrystalline constitution (cause)
Outline

- Deep Learning: the Basics
- Materials Discovery: the Basics
- Deep CNN for EBSD Indexing
- Advanced Software and Hardware in Use
Electron backscatter diffraction (EBSD) is a standard technique detecting certain microstructure characteristics on the surfaces of polycrystals.

Traditional approach to EBSD indexing is pattern matching.

- Store every distinct pattern-angle pair
- build a pattern-angle dictionary
- When a new pattern is observed, it is looked up in the dictionary and the orientation of its 1-Nearest Neighbor (1-NN) is returned

High cost at prediction time!
**EBSD Patterns**

**Electron backscatter diffraction** (EBSD) is a standard technique detecting certain microstructure characteristics on the surfaces of polycrystals.

Traditional approach to EBSD indexing is pattern matching.

- Store every distinct pattern-angle pair
- build a pattern-angle dictionary
- When a new pattern is observed, it is looked up in the dictionary and the orientation of its 1-Nearest Neighbor (1-NN) is returned
Deep Learning Solution

Forward: get predictions

Backward: propagate errors
Loss Function

• Usually in a regression problem the loss to be propagated is the Mean Square Error (MSE) between predicted value and target value.

• However in our problem the target variable is an angle; the periodicity of angular data has to be addressed.

• We designed a special loss function to account for the fact that in angular values 0 is close to 359

\[ L_i(y_i, \hat{y}_i) = \arccos(\cos(\|y_i - \hat{y}_i\|)) \]

predicted angle
true angle
Network Configuration

60 x 60 x 1
68 x 68 x 32
60 x 60 x 32
30 x 30 x 32
30 x 30 x 64
30 x 30 x 64
pool

ϕ
φ
φ_1
φ_2

512
256
3
Prediction Results

On average we are 54% better than state-of-the-art benchmark.

# Time Results

<table>
<thead>
<tr>
<th>Predictor</th>
<th>MAE (eV / atom)</th>
<th>Training time (entire data)</th>
<th>Run time (entire data)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-NN</td>
<td>1.299</td>
<td>0</td>
<td>375 s</td>
</tr>
<tr>
<td>DNN</td>
<td>0.072</td>
<td>7 days</td>
<td>50 s</td>
</tr>
</tbody>
</table>

The **1-NN runtime at testing** is calculated with all test data (30k) processed in one batch, which will produce with the fastest speed with maximum memory consumption.

The **DNN runtime at testing** is calculated with a batch size of 1k to balance the memory usage and time consumption.
A total of 32 Filters learned by the first Conv layer in our CNN from EBSD images.

Black: weight value towards 0 - not important
White: weight value towards 1 - significant
Activation Visualization

Important information retained while noisy background discarded.
Outline

• Deep Learning: the Basics

• Materials Discovery: the Basics

• Deep CNN for EBSD Indexing

• Advanced Software and Hardware in Use
Advanced Software and Hardware

Hardware

- DIGITS DevBox, featuring:
  - 4 TITAN X GPUs with 12GB of memory per GPU
  - 64GB DDR4
  - Core i7-5930K 6 Core 3.5GHz desktop processor
  - 3 x 3TB SATA RAID5
  - 250GB SSD

Software

- CUDA
- cuDNN
- Theano
Conclusion

• We presented the first large-scale, big data application of deep learning in materials discovery


• Advanced software frameworks and hardware infrastructure are adopted.
Materials Discovery: Understanding Polycrystals from Large-Scale Electron Patterns

3rd Workshop on Advances in Software and Hardware (ASH) at BigData
December 5, 2016

Ruoqian Liu, Dipendra Jha, Ankit Agrawal, Wei-keng Liao, Marc De Graef, Alok Choudhary

Thank you!