

DOI: 10.21767/2394-9988.100057

## Automated Data Analysis Strategy for Synchrotron Experiments

**Lifen Yan\***

Department of Computer Engineering, Santa Clara University, Santa Clara, CA, USA

\*Corresponding author: Lifen Yan, Department of Computer Engineering, Santa Clara University, Santa Clara, CA, USA, Tel: +1-530-848-6585; E-mail: lyan@scu.edu

Received date: Jul 10, 2017; Accepted date: Aug 14, 2017; Published date: Aug 21, 2017

Citation: Yan L. Automated Data Analysis Strategy for Synchrotron Experiments. Int J Appl Sci Res Rev. Vol. 4 No. 1:6

Copyright: © 2017 Yan L. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### Editorial

Over the last 30 years, synchrotron-based techniques have had a profound effect on the study of complex materials and on science generally. With the development of femtosecond time-resolved and nanometer space-resolved synchrotron techniques, it has become not uncommon to collect a terabyte of data in a very short amount of time. This poses a daunting task from the perspective of data analysis. Also, a complete understanding of a complex material system often requires a suite of characterization tools that can reveal its elemental, structural, chemical and physical properties at different length-scales and time-scales. Modern scientific studies require researchers to correlate ever-increasing amounts of data and have become an increasingly common bottle-neck to progress.

Complex data analysis takes time, limiting the capacity of researchers to address scientific questions. Currently, the standard workflow for most synchrotron users includes planning and preparing for synchrotron beam time, collecting data over several days, and then returning with significant amounts of data for later analysis. An automated synchrotron data analysis strategy can provide the user with immediate feedback on their measurement results, shorten the latency between measurement and interpretation, and ultimately contribute to greater scientific productivity [1].

Machine learning can be used to automatically detect patterns and help extract meaningful insights from massive amounts of data and is well-suited to image-like and time-series datasets. The predictive power of machine learning has made it a growing method across many scientific domains and machines can be trained to perform the more time-consuming aspects of synchrotron workflow. Autonomous data processing and data fitting for synchrotron studies is needed and presents an alternate path for researchers to increase their own scientific productivity and that of science in general.

Scientists at Brookhaven National Laboratory are working in this direction. In preliminary work, they collaborated with researchers from University of North Carolina at Chapel Hill to explore the use of computer vision methods for organizing, searching and classifying x-ray scattering images [2]. An experimental dataset of 2832 gray-scale x-ray scattering images were processed. They used traditional computer vision

techniques for the classification work, and examined 7 hand-designed features, such as HOG [3] and SIFT [4], which are commonly used in computer vision [2]. They utilized Support Vector Machines (SVMs) [5] to learn visual models for attributes on these images, and demonstrated applications involving attribute annotation and image retrieval. System clusters the images into sets of similar images, which can help scientists to automatically sort their data.

Compared to the traditional approaches of defining hand crafted features, deep learning methods are powerful to learn features automatically [6]. Recent advances in Machine Learning, especially unsupervised feature learning with deep neural networks [7], hold potential to effectively learn basic patterns or features of unlabeled imagery without any supervision. Following this direction, researchers at Stony Brook University applied deep learning methods involving a Convolutional Neural Network (CNN) [8] to automatically recognize x-ray scattering image features in data streams from National Synchrotron Light Source II (NSLS-II) at Brookhaven National Laboratory, and integrated their deep-learning methods into Google Tensor flow to cluster and label the 2-D scattering image patterns [9,10]. Besides the 2832 images in Kiapour et al. [2], they also generated 100,000 synthetic x-ray images to train CNN. Their experiments show that the CNN-based image labeling attains a 10% improvement in the mean average precision over traditional K-mean and Support Vector Machines.

Computer-aided processing of beamline datasets has the potential to alleviate the more time-consuming aspects of experimental workflow. With such an automation system, computer-directed beamline experiments would allow a greater and more efficient exploration of physical parameter spaces. Many x-ray imaging techniques e.g. x-ray scattering, x-ray diffraction, and other x-ray fluorescence microscopies, may potentially benefit as well. And there are lots of existing data which can be used to train machine based systems.

In all, Machine Learning has the potential to become an important component of materials research. It is a new data-processing strategy that can potentially facilitate scientific research, enabling computer-assisted ‘intelligent’ exploration of materials questions, and accelerate scientific discoveries. An automated synchrotron data-analysis system can provide users

with immediate feedback on their measurement results, and the training of machine-learning algorithms can enable an implicit sharing of data between research teams that can enhance the progress and insights for all.

## References

1. Berry M, Potok TE, Balaprakash P, Hoffmann H, Vatsavai R, et al. (2015) Machine learning and understanding for intelligent extreme scale scientific computing and discovery.
2. Kiapour MH, Yager K, Berg AC, Berg TL (2014) Materials discovery: Fine-grained classification of x-ray scattering images. IEEE Winter Conference on Applications of Computer Vision, USA
3. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. Computer Vision and Pattern Recognition (CVPR), USA.
4. Lowe D (1999) Object recognition from local scale-invariant features. Proceeding of the 7th International Conference on Computer Vision (ICCV), USA.
5. Fan RE, Chang KW, Hsieh CJ, Wang XR, Lin CJ (2008) Liblinear: A library for large linear classification. *J Mach Learn Res* 9: 1871-1874.
6. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521: 436-444.
7. <http://deeplearning.stanford.edu/wiki/index.php>
8. Krizhevsky A, Sutskever I, Hinton G (2012) ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, Canada.
9. Wang B, Yager K, Yu D, Hoai M (2017) X-ray scattering image classification using deep learning. Proceedings of Winter Conference on Applications of Computer Vision, USA.
10. Wang B, Guan Z, Yao S, Qin H, Nguyen MH, et al. (2016) Deep learning for analysing synchrotron data streams.