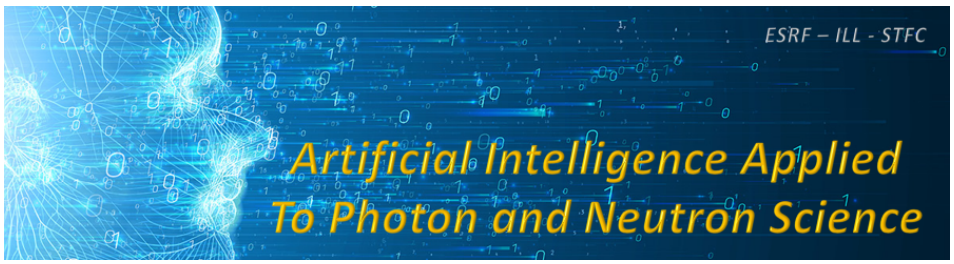
Artificial Intelligence Applied to Photon and Neutron Science

Tuesday 12 November 2019 - Thursday 14 November 2019



Book of Abstracts

Contents

Afternoon 1 / **39**

Introduction to Machine Learning and Deep Neural Networks for scattering science

Corresponding Author(s):

I will give a brief introduction to modern machine learning and deep learning techniques aimed at researchers planning to use them for X-ray and neutron scattering applications. Areas covered will include basic ML terminology and concerns, a quick tour of some probabilistic methods including Gaussian Processes (Kriging), and a discussion of modern neural methods including deep nets, auto-encoders, adversarial training, recursive nets, etc.

Afternoon 1 / **40**

ESRF Overview

Corresponding Author(s):

Afternoon 1 / **41**

Machine learning applications for Small Angle X-ray Scattering data collection and analysis at EMBL-Hamburg

Corresponding Author(s):

Afternoon 2 / **102**

The CAMERA project at Berkeley National Lab

Corresponding Author(s):

Afternoon 2 / **42**

Machine Learning at ILL

Corresponding Author(s):

Recently, by using deep learning methods, computers are able to surpass or come close to matching human performance on image analysis and pattern recognition. This advanced method could also help interpreting data from neutron scattering experiments. Those data contain rich scientific information about structure and dynamics of materials under investigation, and deep learning could help researchers better understand the link between experimental data and materials properties. We applied deep learning techniques to scientific neutron scattering data. This is a complex problem due to the multi-parameter space we have to deal with. We have used a convolutional neural network-based model to evaluate the quality of experimental neutron scattering images, which can be influenced by instrument configuration, sample and sample environment parameters. Sample structure can be

deduced during data collection that can be therefore optimised. The neural network model can predict the experimental parameters to properly setup the instrument and derive the best measurement strategy. This results in a higher quality of data obtained in a shorter time, facilitating data analysis and interpretation.

Afternoon 2 / **43**

Machine learning algorithms for image processing in CryoEM

Corresponding Author(s):

Single particle analysis by Electron Microscopy is a well established technique to analyze the three-dimensional structure of biological macromolecules. The acquired images have a signal-to-noise ratio between 0.1 and 0.01 so that all the image processing steps require to be very robust to extremely high levels of noise. Machine and deep learning algorithms have such characteristics when trained with a sufficiently large amount of data. In this talk we will review the applications of these families of algorithms to the different image processing steps along the image analysis pipeline.

Morning 1 / **67**

Biomedical image reconstruction: From the foundations to deep neural networks

Corresponding Author(s):

Morning 1 / **68**

Machine Learning for improving image quality in tomography

Corresponding Author(s):

In recent years, several imaging fields, including computed
tomography, have benefited from the use of deep learning methods. Nonetheless, successful practical application of these
techniques is often inhibited by the lack of sufficient training data. In this talk, we present several approaches for applying deep neural networks to tomography problems where little or no training data is available. These neural networks can for instance be used to improve reconstruction quality, enabling analysis of more challenging samples than is currently possible. Results will be shown for various types of objects, and practical considerations, such as computational requirements and generalizability, will be discussed.

Morning 1 / **69**

What does artificial intelligence see in 3D protein structures?

Corresponding Author(s):

Structural bioinformatics and structural biology in the last 40 years have been dominated by bottom-up approaches. Specifically, researchers have been trying to construct complex models of macromolecules starting from the first principles. These approaches require many approximations and very often turned out to be rough or even incorrect. For example, many classical methods are based on a dictionary of structural features determined by expert knowledge, such as protein secondary structure, electrostatic estimations, solvent accessibility, etc. However, the reality and underlying physics of proteins is much more complex than our current description of it. Therefore, more progress is needed in this field. Fortunately, deep learning has recently become a very powerful alternative to many classical methods, as it provides a robust machinery for the development of top-down techniques, where one can learn elementary laws from a number of high-level observations. Indeed, it allows constructing models using features and descriptors of raw input data that would be inaccessible otherwise. We have recently studied recurrent structural patterns in protein structures recognized by a deep neural network. We demonstrated that neural networks can learn a vast amount of chemo-structural features with only a very little amount of human supervision.

Our architecture learns atomic, amino acid, and also higher level molecular descriptors. Some of them are rather complex, but well understood from the biophysical point of view. These include atom partial charges, atom chemical elements, properties of amino acids, protein secondary structure, and atom solvent exposure. We also demonstrate that our network architecture learns novel structural features. For example, we discovered a structural pattern consisting of an arginine side-chain buried in a beta-sheet. Another pattern is a spatially proximate alanine and leucine residues located on the consecutive turns of an alpha helix. Overall, our study demonstrates the power of deep learning in the representation of protein structure. It provides rich information about atom and amino acid properties and also suggests novel structural features that can be used in future computational methods.

Morning 2 / 70

Data-driven Materials Discovery for Functional Applications

Corresponding Author(s):

This talk will showcase the use of artificial intelligence (natural language processing, optical character recognition, and machine learning) to auto-generate materials databases for application to areas of interest for neutron science (and the wider materials science community). Specifically, software tools that auto-extract and autonomously analyse materials characterization data will be presented, using case studies for their demonstration, taken from the areas of magnetism and materials for energy.

Morning 2 / 71

Applications of Artificial Neural Networks in Electron Microscopy

Corresponding Author(s):

Morning 2 / 72

Regular Speaker

Morning 2 / **73**

Regular Speaker

Afternoon 1 / **74**

Machine learning for an XFEL accelerator

Corresponding Author(s):

X-ray Free Electron Lasers (XFELs) are among the most complex modern accelerator facilities. With large parameter spaces, highly non-linear behavior, and large data rates, there are expanding opportunities to apply machine learning to XFEL operation and design. In this talk I will give an overview of the challenges, and will cover several applications of machine learning, including online optimization, surrogate modeling, computer vision, and multiplex data acquisition

Afternoon 1 / **75**

Machine Learning at Argonne National Lab

Corresponding Author(s):

Afternoon 1 / **76**

Scientific Machine Learning Benchmarks

Corresponding Author(s):

The use of artificial intelligence (AI) technologies, and of deep learning neural networks in particular, is already having a major impact on many aspects of our lives. The challenge for scientists is to explore how these technologies could have a similar impact for scientific discovery. Already Google DeepMind's AlphaFold tool has achieved some impressive results for protein folding predictions.

The Scientific Machine Learning (SciML) Group at the Rutherford Appleton Laboratory in the UK is focused on applying a range of AI technologies, including deep learning, to scientific data generated by the large-scale scientific experimental facilities on the Harwell site. The SciML group is therefore working with researchers at the Diamond Light Source, the cryo-Electron Microscopy facility, the ISIS Neutron and Muon Source, the Central Laser Facility (CLF) and the Centre for Environmental Data Analysis (CEDA).

This talk will share some initial experiences of our 'AI for Science' explorations in collaboration with the UK's Alan Turing Institute. The talk will then focus on the development of a AI-centric benchmark suite specialised for scientific applications. We believe that such a benchmark suite will help scientists map out the suitability of different AI techniques for different scientific problems. Research into the robustness of results from machine learning technologies as well as on uncertainty quantification will be important to gain confidence into the reliability and understandability of these techniques.

Afternoon 2 / **77**

Machine learning to accelerate materials discovery, modeling, and experiment at ORNL

Corresponding Author(s):

The combination of high-performance computing and machine learning provides routes to address the most challenging problems in materials and experiment. Among the foremost challenges currently is accelerating the discovery and understanding of quantum materials. Crucial to this is the prediction of emergent properties in materials and the interpretation of experimental data. Magnetism is a key route to realizing quantum materials such as magnetic Weyl semimetals, potential building blocks for topological quantum computers. Here, I will cover the challenges in predicting magnetic materials, and the approach to validation and experiment. An essential part of this is to use machine learning to model high dimensional interaction spaces to identify novel quantum phases and to train machines to interpret neutron scattering data. Autoencoders are being used to denoise, remove background, and interpret diffuse neutron diffraction, single crystal inelastic, and powder scattering data. Most importantly, machine learning provides ways to handle the inherent complexity of states and interactions in materials that have been previously beyond reach.

Afternoon 2 / **78**

Machine learning to accelerate materials discovery, modeling, and experiment at DIAMOND

Corresponding Author(s):

The combination of high-performance computing and machine learning provides routes to address the most challenging problems in materials and experiment. Among the foremost challenges currently is accelerating the discovery and understanding of quantum materials. Crucial to this is the prediction of emergent properties in materials and the interpretation of experimental data. Magnetism is a key route to realizing quantum materials such as magnetic Weyl semimetals, potential building blocks for topological quantum computers. Here, I will cover the challenges in predicting magnetic materials, and the approach to validation and experiment. An essential part of this is to use machine learning to model high dimensional interaction spaces to identify novel quantum phases and to train machines to interpret neutron scattering data. Autoencoders are being used to denoise, remove background, and interpret diffuse neutron diffraction, single crystal inelastic, and powder scattering data. Most importantly, machine learning provides ways to handle the inherent complexity of states and interactions in materials that have been previously beyond reach.

Afternoon 2 / **79**

Regular Speaker

Afternoon 2 / **80**

Regular Speaker

Morning 1 / 81

Deriving the big picture from huge spatial datasets: How to make a little training data go a long way

Corresponding Author(s):

Analysis of massive spatial sensor datasets has been revolutionized by the advent of neural network methods. This class of methods has enabled identification and classification of spatio-temporal patterns and objects at multiple scales. Neural network methods need to be trained; the disparity between the very limited human ability to generate training data and the overwhelming training data requirements associated with massive dataset analyses has created a need to develop novel training technologies. I will describe the problem and discuss a variety of methods developed by our group and others to tackle this problem. I will give examples of these methods in the context of large scale microscopy and satellite analyses and discuss how these ideas should apply to broader sets of massive photon and neutron sensor analysis challenges.

Morning 1 / 82

Deep Generative Models for detector simulation

Corresponding Author(s):

The High Energy Physics (HEP) community has a long tradition of using Machine Learning methods to solve tasks related, mostly, to the selection of interesting events over the overwhelming background produced at colliders. In recent years, several studies, in different fields of science, industry and society, have demonstrated the benefit of using Deep Learning (DL) to solve typical tasks related to data analysis. Building on these examples, many HEP experiments are now working on integrating DL into their workflows for different applications: from data quality monitoring, to real-time selection, to simulation.
In particular, Monte Carlo simulation is expected to represent one of the major challenges, in terms of computing resources, for the High Luminosity LHC upgrade and alternative fast simulation solutions will be required.
In this talk, I will present several studies on the use of Generative Models as potential alternatives to classical simulation. Initial results are very promising: different levels of agreement to Monte Carlo have been reached. Most studies are now beyond the initial prototyping stage, and face new challenges related to detailed performance assessment, optimisation, computing resources and integration in the simulation framework.

Morning 1 / 83

Regular Speaker

Morning 2 / 85

Machine learning accelerated analysis of materials data: The Smart facility

Corresponding Author(s):

If data is oil, then national facilities are vast rich fields - producing terabytes per day. However the great majority of this data is ultimately lost completely. In this talk I will look at how machine learning can allow us to exploit more of this data and extract information. I will present some of the methods that the SciML team at Rutherford Appleton Laboratory is using to accelerate the analysis of materials data. The activities I will cover include (i) Construction of database of inelastic neutron scattering data in association with Oak Ridge National Laboratory and the methods that we are developing based on that data to clean and interpret experimental data. (ii) Work with Diamond Light Source to analyse experimental data from diffuse multiple scattering experiments on piezoelectric materials. (iii) Work with industrial and academic partners to analyse images from X-ray imaging and electron microscopy - applying state of the art methods to clean and classify data. These methods represent part of our program to accelerate the conversion of data to knowledge by including at all stages of experiments at national facilities, maximising outputs and enabling a truly Smart laboratory.

Morning 2 / **103**

Regular Speaker

Morning 1 / **84**

Regular Speaker