Artificial Intelligence Applied to Photon and Neutron Science
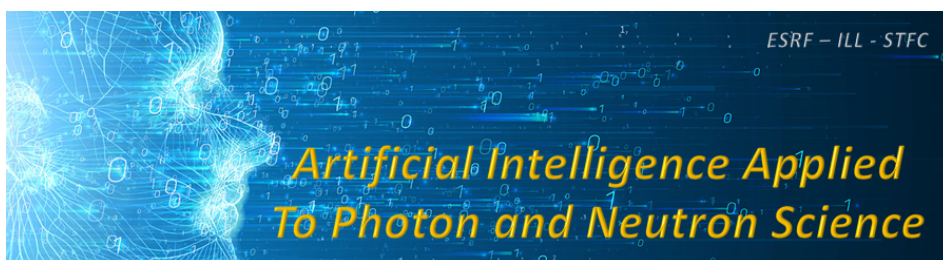
Tuesday, 12 November 2019 - Thursday, 14 November 2019

Book of Abstracts

Contents

Afternoon 1 / 125

Welcome from ESRF Director

Afternoon 1 / 126

Welcome from ILL Director

Afternoon 1 / 39

Introduction to Machine Learning and Deep Neural Networks for scattering science

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

Improving data analysis using artificial intelligence: an ESRF perspective

The ESRF will soon restart after its Extremely Brilliant Source upgrade, which will provide two orders of magnitude improved photon flux for many experiments, and will also come with several new beamlines producing high-throughput data, from macromolecular crystallography to large volume tomography. This creates many challenges in terms of data handling, both from the point of view of the facility (the 'data deluge') and for users who are more focused on practical results on materials rather than methodology details.

We will discuss a number of domains where artificial intelligence could be used to improve the workflow from the experiment to quantitative analysis, including:
- reduction of data by detecting relevant datasets (e.g. serial experiments)
- feature recognition in various techniques (imaging, spectroscopy, diffraction)
- improved (faster) algorithms for data inversion (e.g. coherent scattering experiments)
- more unsupervised/automated workflows for standard experiments, to broaden the user community, including industrial applications

Afternoon 1 / 41

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

In recent years, machine learning and artificial intelligence rapidly gained popularity in many fields

of industry and research, particularly as a tool capable of extracting information from amounts of data often too large to analyze manually. Small Angle X-Ray Scattering (SAXS) of biological macromolecules in solution is routinely being used to evaluate the structural parameters and low resolution shapes of the specimen under study. Here, the application of machine-learning methods seems to be a natural extension, maybe even evolution, of established analysis methods.

At EMBL in Hamburg, two applications of machine learning methods have previously been developed: firstly, based on 450.000 scattering patterns predicted from geometrical objects with uniform density (BODIES; Konarev et al., 2003) and 550.000 scattering patterns from random chains (EOM; Tria et al., 2015), a k-Nearest-Neighbor (kNN) learner is utilized to reliably distinguish between compact objects with or without cavities, extended and flat objects, random chains, as well as "unrecognizable data", i.e. anything not suitable for the other categories. Secondly, based on 150.000 atomic structures from the PDB (Berman et al. 2000) and their calculated scattering patterns (CRYSOL; Svergun et al., 1995), a similar kNN learner has been to evaluate the maximum dimension (Dmax) and Molecular Weight (MW) from the input data. In both cases the same data transformation into a reduced feature space is applied: the theoretical scattering data, which may be provided on any scale and with any angular spacing is transformed to dimensionless Kratky scale (Durand et al., 2010) and subsequently integrated up to sRg = 3, 4 and 5, respectively. The results of this integration are used as input features for learning. To evaluate the performance, the available data has been randomly split into training and cross-validation data sets. Performance of shape classification is rated at 99% for both F1-score an Matthews Correlation Coefficient (Matthews, 1975), with all categories exceeding 90% of one-vs-all classification accuracy. Further, about 90% of continuous Dmax and MW estimates are within 10% of the expected value obtained from the corresponding databank-entry (Franke, 2018).

In addition, recent work aims to employ classical neural networks and/or deep-learning to more challenging tasks. In particular, extensions of the previous work to determine the radius of gyration (Rg, Dmax and MW) from the experimental data directly. Even further, early work suggests that it is possible to develop deep-learning networks to retrieve feasible approximations of the inverse Fourier Transform of the experimental data. A preview version of this has been implemented as a public web, providing the possibility to inspect and download the results (https://dara.embl-hamburg.de/gnnom.php).

Afternoon 2 / 102

DOE's Center for Advanced Mathematics for Energy Research Applications (CAMERA): Artificial Intelligence, Machine Learning, and Experimental Facilities: Present and Future

The Center for Advanced Mathematics for Energy Research Applications (CAMERA) is a US Department of
Energy-wide institute focused on building and deploying mathematical algorithms to accelerate our ability to understand data coming out of synchrotron light sources. CAMERA, consisting of interdisciplinary
teams of applied mathematicians, statisticians, signal processors, computer scientists, software engineers,
physicists, chemists, biologists, and beam line scientists, works in a variety of areas, including ptychography, tomography, SAXS/WAXS/GISAXS, single particle imaging, fluctuation scattering, XPCS,
real-time experimental feedback, computer vision and image processing, machine learning for materials
analysis and image extraction, and real-time autonomous optimized experiments.

This talk with discuss some of those topics, as well as try to address the questions:

(1) "What will experimental facilities look like in the future?";
(2) What role will AI and machine learning play in accelerating our abilities to maximize experimental facilities"; and
(3) "What will we have to build to make this happen?"

Afternoon 2 / 42

Machine Learning at ILL

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

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.

Afternoon 2 / 122

Machine learning and artificial intelligence in MX

In recent years, large-scale facilities such as synchrotrons have encountered a steep increase in demand for computational resources. This is mainly due to very large data rates and volumes and a growing interest by users in real-time feedback during an experiment. Additionally, using X-ray crystallography has gradually become a standard laboratory method rather than a scientific discipline and users now are rarely highly trained experts in the field. Therefore, providing a high-performance computing environment in combination with machine learning applications offers a great opportunity to support and assist users at various key stages during their diffraction experiment.
A series of classification tools has been developed to assist beamline users during decision making when collecting and assessing their X-ray diffraction data. A database, METRIX, was created to hold

a default set of training data which is used to train standard base classifiers located at key decision-making steps based on data processing and model statistics. They help the user assess the chances for experimental phasing success and whether a resulting electron density map is of sufficient quality to attempt model building.

The current focus is on data processing and experimental phasing only, but implementations are underway to include data collection and beamline details as well as molecular replacement. Furthermore, including results of other prediction tools, e.g. for secondary structure and contacts, has been considered.

Afternoon 2 / 123

Convolutional neural networks for DESY photon science

We are exploiting possible applications of artificial intelligence at the German electron synchrotron (DESY) in Hamburg, in particular in the field of photon science. Our current focus is on the use of convolutional neural networks applied to 2D and 3D image analysis for life science.

We will present successful applied semantic segmentation of volumetric 3D synchrotron radiation micro-computed tomography (SRµCT) data with a U-Net. We have trained a convolutional neural network to segment biodegenerable bone implants (screws) and degeneration products from bone and background. The results obtained significantly outperform the previously used semi-automatic segmentation procedure in terms of accuracy and has successfully been applied to more than 100 rather heterogeneous datasets. Remarkably the performance of the U-Net segmentation is considerably better than the experts segmentation that has been used for training.

In addition our ongoing work tackling for instance segmentation (SRµCT) in the context of material science and object detection and classification for cryo electron tomography will be introduced. The machine learning efforts at DESY-IT also include the development of a classification/filter method for XFEL SFX diffraction data.

Morning 1 / 67

Biomedical image reconstruction: From the foundations to deep neural networks

We present a unified overview of biomedical image reconstruction via direct, variational, and learning-based methods. We start with a review of linear reconstruction methods (first generation) that typically involve some form of back-propagation (CT or PET) and/or the fast Fourier transform (in the case of MRI). We then move on to sparsity-promoting reconstructions algorithms supported by the theory of compressed sensing. These are the most popular representative of the variational methods, which are typically iterative (second generation). Finally, we describe the most recent techniques based on convolutional neural networks (third generation), which constitute the current frontier.

While the second and third generation methods are aimed at reducing the radiation dose (faster imaging), they also have the ability to improve image quality under normal acquisition conditions (full exposure). The third generation methods yield the best performance (SNR), but they are not as robust and well understood as the second generation ones. The first-generation methods (efficient linear solver) retain their importance as a critical module of the second and third generation methods.

Morning 1 / 68

Machine Learning for improving image quality in tomography

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?

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 macro-molecules 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 struc-tures 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 struc-ture, 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

This talk will showcase the use of artificial intelligence (natural language processing, optical charac-ter 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

Being charged particles, electrons have a more than 4 orders of magnitude stronger interaction with matter than X-rays or neutrons and may be focused into a spot with less than half an Angstrom in diameter. This makes electron microscopes (EM) very versatile tools for high-resolution imaging, but also diffraction and spectroscopy from very small volumes. The strong interaction with matter, however, comes at the cost of having to account for a complex scattering mechanism when aiming for quantitative comparison of experiment and simulation. In this talk I will present our own work of recovering the 3D structure of the scattering object from multiple electron scattering using deep artificial neural network (ANN) architectures, as well as the application of a multi-scale convolutional neural network for generic image reconstruction tasks.

Morning 2 / 72

Deep learning for classifying and sorting diffraction images

Intense pulses from free-electron lasers and high-harmonic-generation sources enable diffractive imaging of individual nanoparticles in free-flight with a single short-wavelength laser shot. The size of the data sets necessary for successful structure determination, often up to several million diffraction patterns, represents a significant problem for data analysis. Usually, hand-made algorithms are developed to approximate particular features within the data with the goal to reduce the dataset size and to filter out irrelevant images, but such approaches do not generalize well to other datasets and are very time-consuming.

Recently, we have shown in [1,2] that deep neural networks can be used to classify large amounts of diffraction data if a smaller subset of the data is labeled by a researcher and used for the training of the network. We found that deep neural networks significantly outperform previous attempts for sorting and classifying complex diffraction patterns and can improve post-processing of large amounts of experimental coherent diffraction imaging data.

Going beyond this approach, we here present first results using unsupervised deep neural networks. A combination of training a variant of the variational auto-encoder factorVAE [3] and a traditional clustering algorithm, namely Hierarchical Density-Based Spatial Clustering of Applications with Noise [4] are utilized. This approach allows us to find characteristic classes of patterns within a data set without any a priori knowledge about the recorded data. Our ultimate goal is to reduce the amount of time, a researcher has to spend on sorting and classifying datasets, to an absolute minimum. Our unsupervised approach will play very well as a pre-sorting step with our already published supervised routine. Datasets with several hundreds of GByte should be sortable and classifiable within few days time instead of multiple weeks. In addition, the unsupervised routine is applicable to online-analysis during experiments. A trained VAE can pre-sort diffraction images while recoding them, making it a valuable asset during experiments.

[1] Zimmermann et al. Phys. Rev. E 99, 063309 (2019)
[2] Langbehn et al. Phys. Rev. Lett. 121, 255301 (2018)
[3] Kim, H. & Mnih, A. arXiv:1802.05983 (2018)
[4] McInnes L, Healy J. (ICDMW), IEEE, pp 33-42. 2017

Morning 2 / 73

PyFitit: the software for quantitative analysis of XANES spectra using machine-learning algorithms

X-ray absorption near-edge spectroscopy (XANES) is becoming an extremely popular tool for material science thanks to the development of new synchrotron radiation light sources. It provides information about charge state and local geometry around atoms of interest in operando and extreme conditions. However, in contrast to X-ray diffraction, a quantitative analysis of XANES spectra is rarely performed in the research papers. The reason must be found in the larger amount of time required for calculation of a single spectrum compared to a diffractogram. For such time-consuming calculations, in the space of several structural parameters, we developed an interpolation approach proposed originally by Smolentsev et al. The current version of this software, named PyFitIt, is a major upgrade version of FitIt and it is based on machine learning algorithms. We have chosen Jupyter Notebook framework to be friendly for users and at the same time being available for remastering. The analytical work is divided in two steps. First, the series of experimental spectra are analysed statistically and decomposed into principal components. Second, pure spectral profiles, recovered by principal components, are fitted by theoretical interpolated spectra. We implemented different schemes of choice of nodes for approximation and learning algorithms including Gradient Boosting of Random Trees, Radial Basis Functions and Neural Networks. The fitting procedure can be performed both for a XANES spectrum or for a difference spectrum, thus minimizing the systematic errors of theoretical simulations. The problem of several local minima is addressed in the framework of direct and indirect approaches.

Afternoon 1 / 74

Machine learning for an XFEL accelerator

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

We overview artificial intelligence and machine learning for science activities at Argonne National Laboratory. We particularly emphasize cross-laboratory efforts involving the Advanced Photon Source. These include advances in the confluence of high-performance computing and photon sciences; large-scale reconstruction; new algorithms for automating error-correction and experimental design; and improved control and experimental steering.

Afternoon 1 / 76

Scientific Machine Learning Benchmarks

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 for accelerating understanding from Neutron Scattering Data

In many cases obtaining results from a neutron scattering measurement is solving an inverse problem. Often this problem is solved by time consuming random guessing. However there may be hidden relationships between the model and the data that can be identified by machine learning to accelerate the solution process. A particularly powerful case has been demonstrated for systems with classical magnetic Hamiltonians. Here, using the computational resources of the Oak Ridge Leadership Computing Facility, a network has been trained off simulations of the resulting scattering from Hamiltonians with multiple exchange parameters. Then the network is applied to 3D magnetic diffuse scattering or 4D magnetic spectroscopy data sets acquired from the instruments at ORNL. This network has quickly identified the regions in the vast parameter space requiring greater scrutiny. When additional data is added, (in this case from specific heat measurements) the region of parameter space is further reduced. Similar inverse problem solutions that are less computationally intense, have been demonstrated with SANS and Reflectometery data. More generally systems with itinerant moments can be computationally intensive to model, however using machine learning to search for connections inside the computational model can allow for quicker evaluation of the S(Q,w) and may also open other routes to theoretical understanding of the model systems.

Afternoon 2 / 78

Machine Learning at Diamond Light Source: Past, Present and Future

Around 8 years ago, the B24 beamline at Diamond Light Source identified an issue with its potential data collections, that we didn't have an immediate answer for. The issue was that although the planned beamline could collect a Cryo Soft X-ray Tomography image of sub cellular structure with amazing resolution in a matter of minutes, due to the complexities of radiation damage and acquisition geometry, the data was noisy with several significant artefacts. These issues manifested themselves when trying to derive quantitative information from the datasets, which generally re-

quires the process of semantic segmentation. Unfortunately, at the time, this process could take an expert researcher weeks' to achieve, so without a significant number of PhD and postdocs manually annotating datasets, the full potential of the beamline may not be achieved. Given this complex problem, Diamond looked to the relevant UK academic communities for help. Within a couple of years a joint PhD student with Nottingham Universities Computer Vision Lab, who had experience with segmentation of complex biological volumes of root structures, was appointed. They immediately researched the problem, and very quickly identified that standard algorithmic and modelling approaches were not going to work, but that a Machine Learning approach may offer the best chance of success.

Now the resulting software is in widespread use around Diamond, we have new PhD students looking into Deep Learning for image segmentation, and we are starting a 3-year stretch of funding to further expand the approach using the Citizen Science Platform Zooniverse. In complex biological segmentation, the future seems bright in the area of Machine Learning, and although we have not yet managed to remove the human in the loop, we have achieved a five-fold reduction in the amount of time they have to spend manually working on the problem. These successes, along with an expanding presence on campus and globally though the general media and publications, have also drawn a much larger community at Diamond. These researchers are looking to harness these methodologies in many diverse ways, to analyse, predict and accelerate the work they do on a day to day basis, for example the new DIAD (Dual Imaging and Diffraction) beamline is investing heavily in ML to help guide users to areas of interest in samples which are changing over time. In addition, we are starting to see third party tools that make use of Machine Learning, becoming easily available and simple to integrate, allowing beamlines to easily tackle complex issues such as crystal finding in Macro molecular Crystallography and particle picking in cryo Electron Microscopy Single Particle Analysis.

Diamond Light Source currently collects around 5 Peta-bytes of Data every year, a value which is increasing at an alarming rate. To continue to publish the work conducted at such a facility, involves distilling potentially tens of Tera-bytes of data down to the refined information that can be represented in a tens of Mega-byte journal publication. We already use Machine Learning to help our researchers with this challenging goal, and can only see its use increasing in the future, as the methodologies, applications, and ingenuity of their application increases.

Afternoon 2 / 79

Big Data Science Center at the Shanghai Synchrotron Radiation Facility - The first Superfacility Platform in China

The rapid development of synchrotron facilities has massively increased the speed with which experiments can be performed, while new methods and techniques have increased the amount of raw data collected during each experiment. Traditionally, users collect data during their assigned and limited beamtime and then spend many months analysing them. With the huge increase in data volume, this is no longer possible. As a consequence, only a small fraction of this multidisciplinary and scientifically complex Big Data are fully analysed and, ultimately, used in scientific publications. This is unfortunate because synchrotron beam-time is an expensive resource with respect to money as well as time. Secondly, a lack of appropriate data analysis approach limits the realisation of experiments that generate a large amount of data in a very short period of time, and thirdly, the current lack of automatized data analysis pipelines prevents the fine-tuning of an experimental run during a beamtime, thereby further reducing the efficiency of the beamtime potential usage. This effect, commonly known as the "data deluge", affects the light sources worldwide in several different ways.

In order to address these crucial Big Data challenges, Prof. Alessandro Sepe is leading the deployment of a novel Big Data Science Infrastructure at the Shanghai Synchrotron Radiation Facility (SSRF), Chinese Academy of Sciences, Zhangjiang Laboratory. The Big Data Science Center (BDSC) at SSRF aims, in fact, at fully integrating Synchrotron Big Data with Artificial Intelligence, High Performance Cloud Supercomputing and Real-time remote robotic experiments, in order to create a World-Class User-Friendly Superfacility, aimed at accelerating scientific discoveries and technological advancements. Here, also non-experts can obtain scientifically meaningful results in real-time

from the multidisciplinary science carried-on at Large National Scientific Facilities like SSRF and Zhangjiang Laboratory. This will effectively extend the use of synchrotron facilities to the largest plethora of scientific disciplines ever, thus dramatically increasing the scientific outcome of the Users at Large Facilities like SSRF, while aiming at supporting all the key national scientific needs nationally and internationally.

This seminar will focus on the solution that the BDSC is architecting at SSRF and Zhangjiang Laboratory to address this Big Data deluge issue indeed, which poses a serious challenge to the scientific future of all the Synchrotron, Neutron and XFEL large facilities worldwide.

Afternoon 2 / 80

Deep learning for small angle scattering under grazing incidence

Grazing-incidence small-angle scattering (GISAS) is a well established technique to analyze thin multilayered films containing nano-sized objects. It offers a lot of benefits, however the data analysis is challenging and time-consuming.

Nowadays deep learning is widely applied in various areas of our all-day life. In many image analysis applications it achieves already a human-level performance.

We investigate an opportunity to apply deep learning for GISAS data analysis. The aim is to provide users with a fast and accurate feedback on the sample parameters. Trained deep neural network delivers result in about 200 ms, while manual fitting of a single GISAS pattern takes at least hours.

The focus of the present contribution is an overview of our activities in the field of deep learning for GISAS data analysis: the challenges we meet and the results we have achieved. The following topics will be highlighted as well:

could we benefit from the transfer learning?
which features of the GISAS pattern contribute to the output result?
how do we evaluate the result delivered by the deep neural network?
The achieved results include successful prediction of the rotational distributions of hexagonally arranged nanoparticles from the GISAS pattern.

Afternoon 2 / 124

Towards automated analysis for neutron reflectivity

The current workflow of neutron (and x-ray) reflectivity requires human intervention at almost every stage, from data capture, through data reduction, and into data analysis. While some of these steps necessitate expert knowledge and judgement, being able to remove the ones which do not remains a goal, enabling on experiment feedback for the system being measured, and lowering the barrier to fully analysed data for expert and beginner alike.
We have been investigating the application of machine learning methods to model selection in neutron reflectivity, starting from simple metrics like cosine similarity and working up to complex deep or convolutional neural networks. We present here our journey, and lessons learned, with everything from data representation, to optimising network architectures and the difficulties associated with them all.
Our work enables a generalised reflectivity input (R vs. Q) to be fed into the network, which predicts a scattering length density profile for the sample, then outputs that information into fitting software. Both input priors and output uncertainty are dealt with through a variety of methods. We discuss the implementation of both of these features.

Morning 1 / 81

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

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

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

Deep learning for Synchrotron X-ray Imaging

X-ray imaging scans at today's synchrotron light sources can yield thousands of image frames per second at high resolution. Current and expected data volumes necessitate having reliable, efficient, and fully automated data processing pipelines. Traditional image processes are limited to many cases because they are not robust enough for strong noises and complex patterns. The deep neural network can emulate the way of human to model the image problem and to process the large datasets automatically. I will present my recent progress in applying deep neural networks to synchrotron X-ray imaging problems. I applied and developed three fundamental functions for X-ray imaging problems: image classification, image transformation, and a solver of inverse problems. I tested and applied these basic functions for tomographic rotation axis calibration, diffraction pattern selection, low-dose tomography enhancement, super-resolution X-ray microscopy, X-ray image segmentation, missing angle tomography reconstruction, and phase retrieval. These works proved the advantage of using deep neural networks to improve the speed and accuracy of synchrotron X-ray

imaging.

Morning 2 / 85

Machine learning accelerated analysis of materials data: The Smart facility

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

Machine Learning for Spectroscopy: From Spectra to Atomic Structures

The vast majority of the scientific fields are currently experiencing the machine learning tsunami, as researchers try to exploit their unparallel ability to learn from data to give new insights and make fast predictions of target properties.

Spectroscopy using synchrotron radiation has also seen in recent years an increasing number of applications where machine learning techniques have been used to extract structural descriptors of materials from the experimental data.[1,2] Most of these studies relied on supervised machine learning to train standard neural networks architectures. As the required amount of training data cannot be easily fulfilled using experimental measurements, the training sets are usually constructed using ab initio simulations. Using theoretical simulations allows to generate a large number of spectra corresponding to different structures, but this approach is not without its caveats. Also, the studies examine the application of machine learning to a limited subset of spectroscopic techniques.

In this contribution, I start by giving an overview of the different experiments performed at the ESRF's spectroscopy beamlines. Afterward, I will focus on the recent applications of machine learning in spectroscopy and discuss in more detail the use of theoretical simulations to train the artificial neural networks.

1 J. Timoshenko, D. Lu, Y. Lin, and A.I. Frenkel, J. Phys. Chem. Lett. 8, 5091 (2017).
2 A.A. Guda, S.A. Guda, K.A. Lomachenko, M.A. Soldatov, I.A. Pankin, A.V. Soldatov, L. Braglia, A.L. Bugaev, A. Martini, M. Signorile, E. Groppo, A. Piovano, E. Borfecchia, and C. Lamberti, Catal. Today 336, 3 (2019).

Morning 1 / 84

Regular Speaker