AI & DataScience Engineer

Development of Tools based on Machine/Deep Learning Approaches for Correlative Characterizations for Li-ion Battery

Laboratory: *LRCS, Amiens, France*
Supervisor: *Arnaud Demortière (CR, CNRS)*
Project: *PEPR Battery OpenStorm Project (France2030)*
Email address: *arnaud.demortiere@cnrs.fr*
Telephone: *+33 6 95 76 01 65*
Contract start date: *January-March 2024*
Contract duration: *18 months*
Candidate Profile: *Data Science, Algorithm AI, Software Architect*
Skills: *codePython, AI DeepLearning, data management, DashPlotly, advanced multimodal characterizations*

The recent advancements in characterization techniques, including XANES (X-ray Absorption Near Edge Structure), XRD (X-ray Diffraction), and XCT (X-ray Computed Tomography), have opened up new possibilities in materials analysis. These techniques now offer ultra-fast acquisition, high resolution detection, 4D mapping, and multi-dimensional analyses. With the emergence of big data and AI processing tools, these advancements have paved the way for a new horizon in materials analysis.

The increasing generation of massive data sets from operando experiments and multi-modalities necessitates automated processing and statistical correlations. These tasks are made possible by the development of machine learning (ML) and deep learning (DL) algorithms [1]. ML and DL approaches have demonstrated their effectiveness in various data treatment domains, particularly in image processing. Examples include semantic segmentation [2], super resolution [3], and multimodal correlation [4].

The complex nature of battery systems poses a typical multiscale problem, where multiple characteristic time constants arise from the interaction of transport and material/electrode properties. To obtain a comprehensive understanding of the global dynamical processes in Li-ion batteries, it is essential to conduct multiscale operando experiments. These experiments provide spatiotemporal information on lithium composition, enabling the monitoring of electrochemical reactions and capturing the complete picture of the battery system.

This project aims to utilize various characterization techniques, including image-based methods, spectroscopy, diffraction, and NMR, with different modulations to generate a comprehensive time-series dataset. To achieve this, dedicated electrochemical operando cells, both homemade and commercial, will be employed.

The initial phase of the project will focus on collecting data during electrochemical cycling using identical electrodes for different techniques. To meet the requirements of each technique, specific sample preparations will be carried out using the same reference sample, which could be LFP or NMC. The data acquisition will be conducted under identical electrochemical conditions. This multimodal dataset will serve as the reference dataset, acting as the foundation for our correlative project.

This approach will be repeated for different types of cathode or anode materials, allowing for a comparison of electrochemical information across different operando systems. Through this comparison, differences and similarities can be identified, offering valuable insights into the behavior of various materials in electrochemical processes.

The second phase of the project will involve cleaning, registering, and preparing the raw dynamical dataset. This will be achieved through the utilization of a CNN (Convolutional Neural Network) network for segmentation and classification processes. The output of these processes will serve as input data for correlative approaches. Additionally, ANN (Artificial Neural Network) networks will be applied to the multiscale datasets to identify correlations and patterns among the features at each scale.

To facilitate the connection of patterns across different scales, a super-resolution CNN approach will be optimized. This bridging process will enhance the ability to predict and interpret the complex hierarchical microstructure. Gaining a comprehensive understanding of the multiphase properties of the materials requires the integration of knowledge from various modalities. For example, a CNN-based similarity learning network, combined with non-subsampled shearlet transform (NSST) for image decomposition into low- and high-frequency components, will be used to establish multimodal correlations. This analysis may involve connecting an XRD diffractogram with a metal oxidation state map of the same crystal.

To enhance the interpretability of the dataset, fusion images can be generated by blending multimodal images together. This process increases the ability to extract meaningful information from the dataset, aiding in the overall analysis and understanding of the materials under investigation.

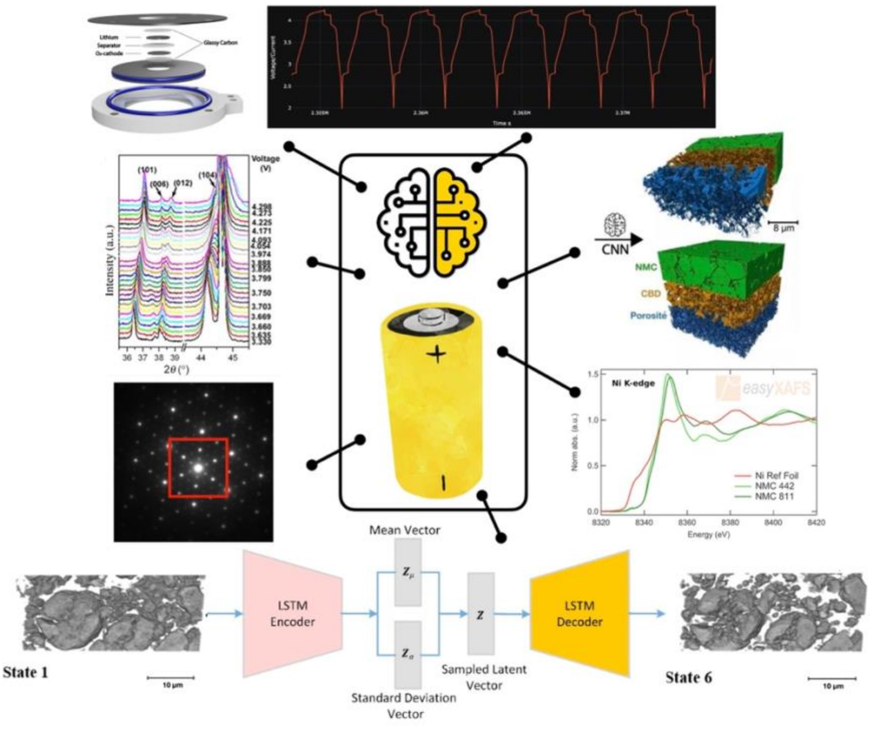

Figure 1. Schematic of the project strategy around the collecting, the cleaning and the processing to correlate all types of datasets on battery multimodal characterizations from electrochemical cycling to structural and chemical properties.

In this project, our aim is to develop a set of tools to assist users in managing large datasets and automating data processing using supervised machine learning and deep learning approaches. We will build specific tools for each characterization technique as outlined below:

Tool 1 - Data Cleaning, Registration, and Segmentation: This tool focuses on processing raw image, spectrum, and diffractogram data to mitigate experimental artifacts such as noise, blur artifacts, rings, and intensity variations, which can impact subsequent steps. Building upon our recent work [5] on unsupervised image quality assessment, this tool will aid in reducing these defaults. Additionally, proper registration, ensuring accurate alignment of images in the time-series, will be performed. The multiphase segmentation will be conducted using our in-house software, SegmentPy (https://segmentpy.readthedocs.io), which utilizes a CNN-based approach.

Tool 2 - Multimodal Correlation Process: To achieve accurate prediction and interpretation of the complex hierarchical microstructure, it is crucial to establish correlations among multiphase properties. Our strategy relies on a CNN-based similarity learning network coupled with the non-subsampled shearlet transform (NSST) for decomposing images, spectra, and diffractograms into low- and high-frequency components. Fusion strategies for imaging will also be employed to enhance the interpretability of the dataset.

Tool 3 - Dynamical Correlation Process: This step involves designing a data-driven strategy for the prepared time-series dataset using an RNN (Recurrent Neural Network) network comprising LSTM (Long Short-Term Memory) or GRU (Gated Recurrent Unit) units. We will optimize the RNN network structure and hyperparameters, following separate and mixing strategies for each sub-scale. We may explore multimodal fusion images, and we will focus on identifying relevant ground truth metrics and developing appropriate loss functions to effectively capture the observed dynamics. Alternative network strategies, such as convolutional-RNN and Prod-RNN, will be explored to incorporate spatiotemporal aspects by combining the advantages of CNN and RNN architectures.

Over the past five years, our team, led by A. Demortière, has made significant progress in image and data processing using deep learning and machine learning algorithms within the Image & Diffraction & Data Science (I&2D) group at LRCS. Our expertise lies in Python coding, handling massive datasets, and training deep learning algorithms such as CNN, VAE (Variational Autoencoder), and LSTM. We have conducted extensive research on multiphase image segmentation, time series analysis, and clustering [6,7].

The selected engineer will engage in extensive collaboration with the laboratories associated with the PEPR batterie OpenStorm project. Their primary focus will involve the development of software tools for efficient data management and the application of AI Deep Learning techniques in analyzing treatment data. This position presents a unique opportunity for the candidate to collaborate with a diverse scientific community working in the fields of battery technology and characterization.

References:

[1] Nelson, J., Misra, S., Yang, Y., Jackson, A., Liu, Y., Wang, H., Toney, M. F. (2012). operando X-ray diffraction and transmission X-ray microscopy of lithium sulfur batteries. *Journal of the American Chemical Society*, *134*(14), 6337-6343.
[2] Hao, S., Zhou, Y., & Guo, Y. (2020). A brief survey on semantic segmentation with deep learning. *Neurocomputing*, *406*, 302-321.
[3] Wang, Z., Chen, J., & Hoi, S. C. (2020). Deep learning for image super-resolution: A survey. *IEEE transactions on pattern analysis and machine intelligence*.
[4] Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., & Ng, A. Y. (2011, January). Multimodal deep learning. In *ICML*.
[5] Burnett, T. L.; Withers, P. J. Completing the Picture through Correlative Characterization. *Nature Materials* 2019, *18* (10), 1041–1049.
[6] Su, Z., Decencière, E., Nguyen, T. T., El-Amiry, K., De Andrade, V., Franco, A. A., Demortière, A. (2022). Artificial neural network approach for multiphase segmentation of battery electrode nano-CT images. *npj Computational Materials*, *8*(1), 1-11. 2022.
[7] Zhang, K., Nguyen, T. T., Su, Z., & Demortière, A. (2022). Self-supervised image quality assessment for X-ray tomographic images of Li-ion battery. *npj Computational Materials*, *8*(1), 194.