Reference ID: 11226820680_Yang

Submission Date and Time: 12/16/2019 3:15:27 PM

Author Names & Affiliations

Submitting author: Judith Yang - University of Pittsburgh

Additional authors: Stephen House, University of Pittsburgh Wissam Saidi, University of Pittsburgh

Contact Email Address (for NSF use only): (hidden)

Research domain(s), discipline(s)/sub-discipline(s)

electron microscopy

Title of Response

Accelerating Infrastructure: In situ Transmisison Electron Microscopy

Abstract

Transmission electron microscopy (TEM) is a powerful and ubiquitous nanoscale characterization tool used in both materials and bio-related scientific fields because of its ability to provide both structural and chemical information down to the Angstrom level. The past two decades have seen two transformational developments in the field of electron microscopy: 1) lens aberration-correctors to

enable sub-Angstrom spatial resolution and 2) in situ electron microscopy that allows for dynamic observations instead of static pre- and post-mortem imaging. However, image analysis is still largely performed by an individual on single micrographs, greatly restricting the amount of scientific information extractable from these large data sets – especially in situ – due to factors ranging from time constraints with human interpretation to inadequacy of commercially available or free-ware image analysis programs. Automated image analysis software, leveraging state of the art machine learning, for rapid and quantitative analysis of these increasing data sets, whether as an individual micrograph or sequence of micrographs (video), has the potential to transform the field of electron microscopy, and thereby transform materials research.

Question 1 (maximum 400 words) – Data-Intensive Research Question(s) and Challenge(s). Describe current or emerging data-intensive/data-driven S&E research challenge(s), providing context in terms of recent research activities and standing questions in the field. NSF is particularly interested in cross-disciplinary challenges that will drive requirements for cross-disciplinary and disciplinary-agnostic data-related CI.

TEM images contain a wealth of information; the challenge is to extract the relevant portions accurately and – particularly for in situ TEM video – in a reasonable time frame. The typical information desired from a given image includes the position, size, and morphology of the features of interest (e.g., NPs, clusters, cells, etc.). From high-resolution images, accurate knowledge of the atomic coordinates in the feature of interest enables the determination of its crystal structure (material identity), orientation, and strain states . For video data, knowing how all these aspects evolve over time (frame-to-frame) and the rates at which these changes occur is necessary for understanding the system dynamics and correlation with measurements of materials performance. Typically, preprocessing of the data is required before analysis to improve signal quality and alignment. Existing, readily available software tools are inadequate for accurately obtaining the above information from in situ TEM datasets. Because of the complex physics involved in atomic-resolution TEM image formation, these images are not directly interpretable, necessitating the use of TEM image simulation to determine the underlying atomic structure. This is especially true at material surfaces and interfaces, which are typically the most scientifically interesting and important parts of a structure. A variety of free and commercial software packages are available for performing these image simulations. However, analyzing even a single image this way is a time-consuming process that requires the involvement of an expert. Analysis of video data consisting of thousands to tens of thousands of individual frames is impractical. A versatile suite of tools needs to be developed to assist the analysis of the wider body of in situ TEM videos, enabling researchers to effectively and efficiently, prepare, process, and analyze data sets collected on a wide range of material systems and instrumentation housed at many public and private research institutions. Accelerating the next historic breakthrough in in situ TEM requires the convergence of forefront computational, statistical, and machine learning tools with these forefront characterization instruments.

Question 2 (maximum 600 words) – Data-Oriented CI Needed to Address the Research Question(s) and Challenge(s). Considering the end-to-end scientific data-to-discovery (workflow) challenges, describe any limitations or absence of existing data-related CI capabilities and services, and/or specific technical and capacity advancements needed in data-related and other CI (e.g., advanced computing, data services, software infrastructure, applications, networking, cybersecurity) that must be addressed to accomplish the research question(s) and challenge(s) identified in Question 1. If possible, please also consider the required end-to-end structural, functional and performance characteristics for such CI services and capabilities. For instance, how can they respond to high levels of data heterogeneity, data integration and interoperability? To what degree can/should they be cross-disciplinary and domain-agnostic? What is required to promote ease of data discovery, publishing and access and delivery?

The critical challenges facing the analysis of in situ TEM video data – and thus where software tools are needed – can be broadly grouped into three categories: mitigating signal obfuscation, feature identification, and the tracking and quantitative analysis of changes in features over time. Mitigating signal obfuscation (blur): The desired signals (the contrast from structural features of interest) of TEM images can be obfuscated by a variety of sources, both experimental and instrumental. These include competing image contrast, such as from shot noise, support materials, or build-up of surface contamination. Another source of obfuscation is the movement of the objects being imaged. Image translation can arise from sample drift during in situ heating, vibrations transmitted to the sample holder, mechanical hysteresis of the holder, or changes in the specimen itself during an in situ reaction. The obfuscating effects can sometimes be addressed with available software tools, but this process tends to be time-consuming, requiring a skilled researcher and many days of effort per data set. Therefore, flexible, automated tools that can detect and account for these effects is needed. Feature identification: Feature identification involves the ability to distinguish and isolate features of interest in a given image/frame. Computationally assisted identification of the features(s) of interest in TEM micrographs and videos is crucial both for decreasing potential issues of human bias and automating the process for analyzing large datasets. The challenge arises primarily from the complex image contrasts that can occur in TEM images. Even when optimal parameters are found that produce the desired segmentation of a given image/frame, they may not produce a good segmentation in other images/frames. Any tools must be flexible enough to account for such variation. Feature identification also includes the precise determination of the atomic positions in the nanostructures being imaged. The atomic-scale contrast is particularly sensitive to changes in microscope parameters, sample thickness, or sample tilt. No existing software tools can account for changes in contrast. Tracking features over time: The challenge of tracking and quantitative analysis of features changes is the difficulty to remove the background and to identify and follow a feature of interest over time – even as its morphology and internal structure may change – and to do so for statistically robust population sizes. Relying on manual data preparation, processing, and analysis greatly restricts the size and number of datasets from which this information can be extracted, limiting the statistical relevance of the observations, and risking inadvertent “cherry-picking.” The ability to track these changes and extract quantitative measurements of their rates at a statistically robust level is critical for correlating TEM observations with bulk reaction kinetics information from ensemble and real-world measurement techniques. One possible solution could be mathematically, deep learning which is the process of estimating a complex function known as a neural networks (NNs) that maps the input (e.g. TEM images) to the desired output (e.g. atomic positions). Deep NNs are simply a sequential application of simpler functions such as linear transformations and pointwise non-linear functions. By applying these simpler functions sequentially, we obtain a much more complex composite function. The choice of the sequence of simple functions to apply to the input data is often known as the architecture of the deep NN, and it is important to select the right architecture for the right problem.

Question 3 (maximum 300 words) – Other considerations. Please discuss any other relevant aspects, such as organization, processes, learning and workforce development, access and sustainability, that need to be addressed; or any other issues more generally that NSF should consider.

Other considerations is the education and training in electron microscopy (for both the theoretical aspects as well as the practical aspects of training to "fly" and data analysis. Close interactions with the microscopy manufacturers are necessary to build and incorporate the software with the microscopes. Another issue is access to these very expensive instruments and the appropriate TEM expert who is also expert in the relevant scientific area of interest.

-- End Submission –