In-Operando Tracking and Prediction of Transition in Material System using LSTM

Pranjal Sahu
Stony Brook University
Stony Brook, New York
psahu@cs.stonybrook.edu

Dantong Yu
New Jersey Institute of Technology
Newark, New Jersey
dantong.yu@njit.edu

Kevin Yager
Brookhaven National Laboratory
Upton, New York
kyager@bnl.gov

Mallesham Dasari
Stony Brook University
Stony Brook, New York
mdasari@cs.stonybrook.edu

Hong Qin
Stony Brook University
Stony Brook, New York
qin@cs.stonybrook.edu

ABSTRACT

The structures of many material systems evolve as they are treated with physical processing. For instance, organic and inorganic crystalline materials frequently coarsen over time as they are thermally treated; with domains (grains) rotating and growing in size. When a material system undergoing the structural transformation is probed using x-ray scattering beams, the peaks in the scattering images will sharpen and intensify, and the scattering rings will become increasingly 'textured'. Accurate identification of the transition frame in advance brings multiple benefits to the NSLS-II in-operando experiments of studying material systems such as minimal beamline damage to samples, reduced energy costs, and the optimal sampling of material properties. In this paper, we formulate the prediction and identification of the structural transition event as a classification problem and apply a novel LSTM model to identify sequences having transition event. The preliminary results of the experiments are encouraging and confirm the viability of the detection and prediction of transition in advance. Our ultimate goal is to deploy such a prediction system in the real-world environment at the selected beamline of NSLS-II for improving the efficiency of the experimental facility.

KEYWORDS

Deep Learning, Autonomous Infrastructure, LSTM, future frame prediction, X ray scattering

ACM Reference Format:
Pranjal Sahu, Dantong Yu, Kevin Yager, Mallesham Dasari, and Hong Qin. 2018. In-Operando Tracking and Prediction of Transition in Material System using LSTM. In *AI-Science'18: Autonomous Infrastructure for Science, June 11, 2018, Tempe, AZ, USA.* ACM, New York, NY, USA, Article 4, 4 pages. https://doi.org/10.1145/3217197.3217204

AI-Science'18, June 11, 2018, Tempe, AZ, USA

ACM ISBN 978-1-4503-5862-0/18/06…$15.00
https://doi.org/10.1145/3217197.3217204

1 INTRODUCTION

X-ray scattering is a technique utilized to probe the physical characteristics of material at the molecular scale [3]. During the process of X-ray scattering, the crystals of the material being examined go through different phases and show specific transformation patterns. Different classes of materials exhibit distinct phases, due to radically different structural order at different temperatures. For instance, liquid crystalline materials are known to undergo a series of phase transitions as a function of temperature, and concentration. Organic self-assembling materials (molecular and polymeric) exhibit a host of re-orientations and phase transformations, including non-equilibrium (irreversible) conversions. Each of these transformations has unique and complex transition patterns: the changes in peak positions, shapes, and intensities and even the appearance and disappearance of peaks when organizational symmetry changes dramatically. These dynamic patterns are observable in x-ray scattering imaging. Figure 1 shows a representative example of this type of transformation process.

Although the phase transformations may be distinct and appear to be evident after their completion; it is often extremely challenging to detect the early signatures of these changes since the beginning stages of conversion involve small, highly disordered grains of the newly appearing morphology (with the corresponding weak and diffuse signal in x-ray scattering). Methods to robustly and rapidly detect transitions at the early stages of reorganization would result in better data sampling and reduction in experimental costs because they would enable automated experiment that identifies the boundaries phase accurately and robustly, tracks the kinetics of material conversion, avoids undesirable transformations, and minimizes redundant sampling of data.

In this paper, we propose an autonomous agent with the objective of predicting the transition stage in an X-ray scattering imaging sequence right before it starts. This will help in dialing up the intensity of beamline and the speed of collecting X-ray observations of the crystal sample during the critical stage where the structural transition event occurs. As a result, the transition prediction will reduce the beamline usage time and minimize the deterioration of crystal sample resulting from unnecessary damage to crystal particles by photons from light sources.

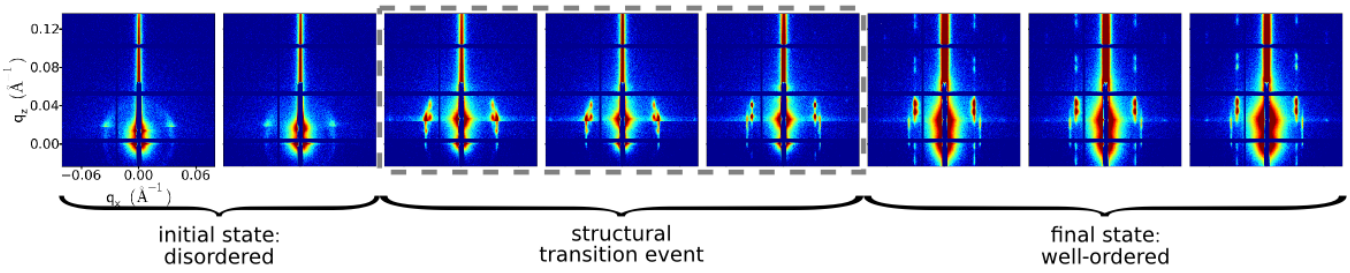

Figure 1: A sample sequence showing eight frames of X-ray In-Operando experiment tracking the structural transition of a crystal sample. A quick transition happens between Frames 3 and 5. The identification of the structural transition event in advance has significant importance since the transition event is rich in information, i.e., obtaining as many samples as possible during this period means better experimental results and more efficient beamline usage.

2 RELATED WORK

Recent advances in deep learning have opened up several new research frontiers that are unthinkable in the past. One of them is the application of predicting future video frames that have a significant resemblance to our problem in scientific experiments. The problem of predicting future frames is defined as follow:

Given a set of n images $[X_1, X_2, ...X_n]$ from a video sequence, the objective is to predict (generate) the next m images $[\hat{X}_{n+1}, \hat{X}_{n+2}, ..., \hat{X}_{n+m}]$ of the given video sequence as closely as possible to the ground truth frames $[X_{n+1}, X_{n+2}.., X_{n+m}]$ between time n and $n + m$. Prior work in this domain of predicting immediate future has focused on getting representations using each frame individually [8]. However, this approach ignores the temporal information among frames. Inspired by the language model for predicting subsequent word embeddings and learning sequence from sequence [6], Srivastava et al. introduced an unsupervised approach in [5] for learning the motion patterns in a video sequence and prediction of the future motion frames. Recently methods using Generative Adversarial Networks have shown promising results in generating the future content [1, 7, 9].

Similar to the related efforts, our objective is to predict the future images in an X-Ray sequence obtained during the in-Operando experiment for tracking the structural transition in the crystals of a material sample. We approach this problem in following two steps:

(1) Identification of sequences having structural transition event.
(2) Prediction of the future frame in the sequence based on historical image frames.

In this paper, we tackle the first problem of the accurate identification of sequences that contain a structural transition event and briefly discuss the future works and specific steps to be taken to obtain a fully functional transition frame prediction system in section 5.

3 METHOD

Taking inspiration from the earlier work in the domain of video sequence classification, we adopt a hybrid deep neural network model that consists of CNN and LSTM for identifying a transition sequence. Training a deep learning model requires a considerable amount of labeled training data that is infeasible to obtain from the real experimental setup. Therefore we adopt a synthetic data generation process to obtain a large number of training samples. In Section 3.1, we describe the training data generation process. And in section 3.2 we describe the CNN-LSTM model that we used for classifying transition sequences.

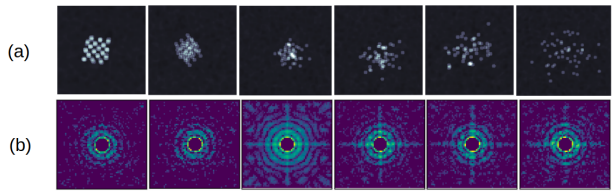

Figure 2: A sequence of artificial real-space crystal and corresponding X-ray scattering data generated for the task of training the CNN-LSTM model. (a) Real-space crystal, (b) X-ray scattering sequence.

3.1 Training Data Generation

We consider three types of crystal structures in our experiment namely BCC (Body-Centered Cubic), FCC (Face Centered Cubic) and SC (Simple Cubic). A crystal structure is selected with some probability, and a 3D box is filled with particles based on the chosen structure (shown in Figure 2). Some random noise of the normal distribution is added to each particle's location in the 3D box. We generate a synthetic sequence of real-space crystal by repeatedly rotating the 3D box by a random angle. To generate the X-ray sequence from the obtained real-space sequence, we first project the particles in the 3D bounding box into a 2D projection, apply a Fast Fourier Transform (FFT) to the projected image in the second step, and in the final step we keep the magnitude image of the spectrum and discard the phase component of the Fourier transformation because detector can only register the amplitude components.

To minimize the computational cost, we generate images with the 64x64 resolution. Since the magnitude images created by FFT have a big range, we take the ten base logarithm of images that are added to a constant value one to prevent the input of zero value. The obtained images are then replicated to form three channels. In a real-life data sampling procedure, the beamline collides with particles and causes the crystal structure to deteriorate gradually.

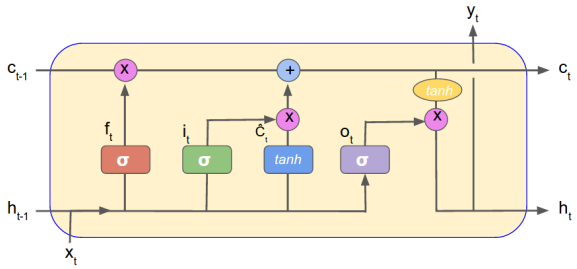

Figure 3: LSTM Architecture [4]. It resolves the long term dependencies problem and helps capturing the temporal information in a sequence. C_t represents cell state, h_t is hidden state, X_t is input and y_t is output.

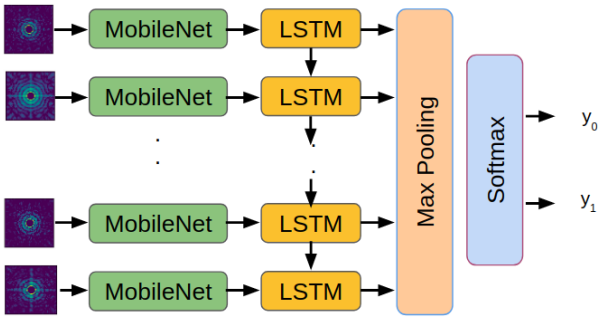

Figure 4: CNN-LSTM model for classification of sequence with or without a transition event.

To simulate this effect, we first translate particles with some random distance in the 3D space to simulate the Brownian motion. It results in the crystal structure gradually losing its shape. The frame at which deterioration starts is termed "transition" frame. In our experiments, we generated synthetic image sequences with the length ranging between 20 and 45. The index of the transition frame is selected randomly. For each generated sequence, we create two data samples, each of which has the length of 10: the first sequence contains transition frame, while the second one does not. After the data generation process, we obtain a total of 20k samples in the training split, with 10K samples containing transition frames and 10K samples without it. Similarly, 4800 samples were obtained in the testing split.

3.2 LSTM based Sequence Classification

We formulate the identification of transition frame(s) as a classification problem. Given a sequence of n frames: $[x_1, x_2...., x_n]$, the LSTM model outputs the likelihood of the sequence containing a transition frame. The transition frame identification is essentially a binary classification problem. Formally this can be interpreted as

$$transitionSequence = \underset{0 \le i \le 1}{\operatorname{argmax}} P(y_i | x_1, x_2, .., x_n, w) \quad (1)$$

where w denotes the weights of the LSTM model that is obtained via the training process. Figure 3 shows a reference architecture of a standard LSTM and the computation process of an LSTM cell is described below:

$$f_t = \sigma(W_f.[h_{t-1}, x_t]i + b_f) \quad (2)$$

$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i) \quad (3)$$

$$\tilde{C}_t = tanh(W_C.[h_{t-1}, x_t] + b_C) \quad (4)$$

$$C_t = f_t * C_{t-1} + u_t * \tilde{C}_t \quad (5)$$

$$o_t = \sigma(W_o.[h_{t-1}, x_t] + b_o) \quad (6)$$

$$h_t = o_t * tanh(C_t) \quad (7)$$

We modified the input component in the LSTM and created a new LSTM architecture (shown in Figure 4) for the purpose of classifying image sequences.

In addition to the temporal relationship among the sequence of images, the X-ray scattering images also have the spatial correlation, a CNN-LSTM approach can better exploit these two context information and therefore is applied in this paper. A Mobilenet CNN [2] pre-trained on ImageNet is used to extract from each frame the spatial features and latent state information that is fed into an LSTM that continues to exploit the temporal information and state transitions contained in the X-ray sequence. The MobileNet CNN uses a depth-wise separable convolution and results in a substantial decrease in the number of learning weights. The weights of the MobileNet CNN are shared across time domain for processing individual frames. The CNN-LSTM network employs a temporal max-pooling layer to aggregate the outputs at each time step across the entire sequence and feeds the pooling result into a softmax classifier that outputs the probability of the sequence having a transition frame. The model comprises around 11 million trainable parameters in total. We trained the network using the back-propagation algorithm and the Adam optimizer. Because we attempt to fine-tuning the pre-trained MobileNet CNN with our X-ray images in a transfer learning configuration, we use a low learning rate of 0.0001. Training batch size is set to 10.

4 RESULTS

In this section, we quantify the performance of the CNN-LSTM classification model. We calculate the Area Under the ROC curve (AUC) as the performance metric for the binary sequence classification problem where the sequence having a transition event belongs to the positive class (label "1"). The ROC curve on the holdout testing data shown in Figure 6 demonstrates that the CNN-LSTM model can classify the sequences with high accuracy as evident from the achieved AUC value of 0.97. Figure 5 shows some sample sequences and their prediction results in the test split. The model is able to capture the transitional event when the changes are prominent. Some sequences have very subtle variations during the transitional event that the model cannot detect and result in a false negative because low-resolution images (64x64) are being used in our current model. We expect that an improved model for the images of a higher resolution can reduce the false positives and generate better performance.

5 FUTURE WORK

In the future work, we will design a model that can also identify and predict the particular transition frame given the past sequence of X-ray image frames. To achieve this, we plan to adopt a multi-task

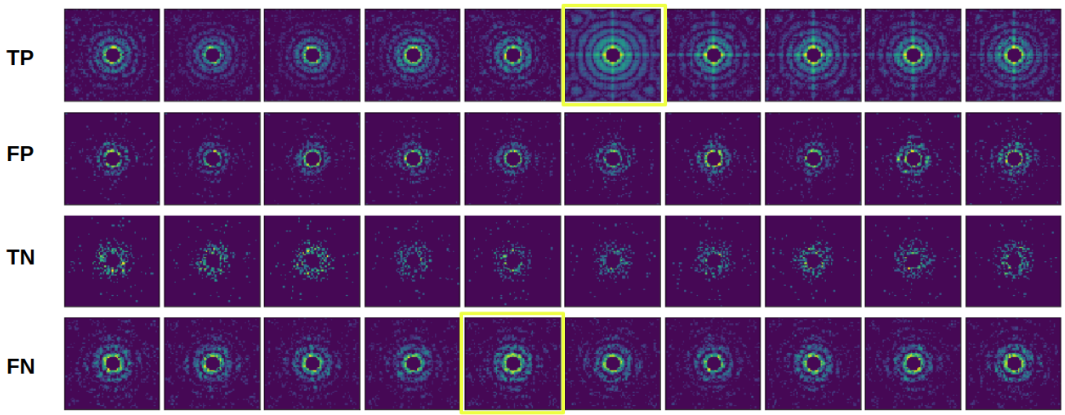

Figure 5: We show four sequence samples in the test split and their classification results, each of which is from True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN). Some sequences have very minute changes that the model makes a false prediction.

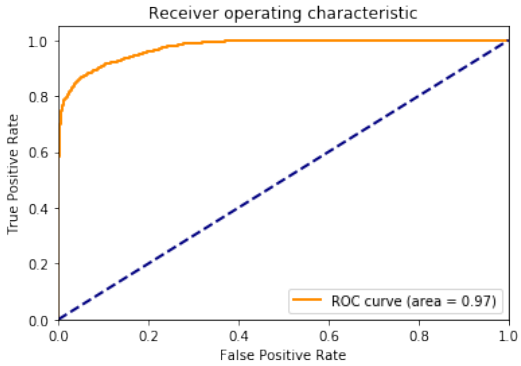

Figure 6: ROC curve for Transition Sequence classification.

learning approach where the objectives will be to construct the future frames as well as identify the transition frame. This will be similar to an encoder-decoder model with two branches, one for classification and other for predicting (constructing) future frames.

We also aim to solve a similar task of using an autonomous agent to explore the state space. For example, a multi-layered block copolymer material begins in a non-equilibrium and kinetically trapped state. Such a material can be annealed in a wide variety of pathways: the combination of different temperature and time histories represents a large and complex state space. Complex non-equilibrium morphologies exist within the space, but are transient, occupying a small region of state space. Those conventional searches are ill-suited to the challenge that reliably generates or even finds such states. A computer-guided autonomous agent can instead select processing pathways during experimentation. The agent-based approach involves both searches for target states (informed by coupling to physical models) and avoidance of transitions to undesired states (by refining the materials model in real time).

6 CONCLUSIONS

In conclusion, we propose a binary classification approach (LSTM-CNN) to predict whether a sequence of experiment images contains a transient event and to enable the possible automation to steer the experiment to focus on the transient event with increased luminosity and probing frequency and skip the static stage. The experiment automation will improve the efficacy of beamline and minimize the sample damage by X-ray. The LSTM-CNN is an approach based on the deep neural network. It considers both spatial and temporal state information and detects the existence of a transient event based on the history of transformation indicating whether the crystal is about to enter the structural transition phase. The preliminary results of sequence classification shows that such a approach is feasible. We will evaluate whether the LSTM-CNN can be extended to explore the state space of a material in an automated fashion.

REFERENCES

[1] Xi Chen, Yan Duan, Rein Houthooft, John Schulman, Ilya Sutskever, and Pieter Abbeel. 2016. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In *Advances in Neural Information Processing Systems*. 2172–2180.

[2] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. 2017. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861* (2017).

[3] Harold P Klug and Leroy E Alexander. 1974. X-ray diffraction procedures: for polycrystalline and amorphous materials. *X-Ray Diffraction Procedures: For Polycrystalline and Amorphous Materials, 2nd Edition, by Harold P. Klug, Leroy E. Alexander, pp. 992. ISBN 0-471-49369-4. Wiley-VCH, May 1974.* (1974), 992.

[4] Chris Olah. 2015. Understanding LSTM Networks. (2015).

[5] Nitish Srivastava, Elman Mansimov, and Ruslan Salakhudinov. 2015. Unsupervised learning of video representations using lstms. In *International conference on machine learning*. 843–852.

[6] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*. 3104–3112.

[7] Sergey Tulyakov, Ming-Yu Liu, Xiaodong Yang, and Jan Kautz. 2017. Mocogan: Decomposing motion and content for video generation. *arXiv preprint arXiv:1707.04993* (2017).

[8] Carl Vondrick, Hamed Pirsiavash, and Antonio Torralba. 2016. Anticipating visual representations from unlabeled video. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 98–106.

[9] Carl Vondrick and Antonio Torralba. 2017. Generating the future with adversarial transformers. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.